Mathematical modelling of metabolism and enzyme-level regulation to understand and infer functional regulators of metabolism

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

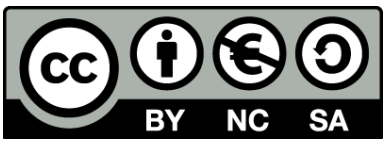
des Fachbereichs Biologie der Philipps-Universität Marburg

vorgelegt von

Herrn Niklas Farke

aus Warstein

Marburg, 2022



Die Arbeit zur vorliegenden Dissertation wurde von Oktober 2017 bis September 2020 an der Philipps-Universität Marburg und von Oktober 2020 bis September 2022 an der Universität Tübingen unter der Betreuung von Herrn Professor Dr. Hannes Link angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation am _____ angenommen.

Erstgutachter: Prof. Dr. Hannes Link

Zweitgutachter: Prof. Dr. Lennart Randau

Weitere Mitglieder der Prüfungskommission: Prof. Dr. Michael Bölker

Prof. Dr. Peter Lenz

Tag der Disputation: 25.11.2022

I Eigenständigkeitserklärung

Hiermit erkläre ich, dass die vorgelegte Dissertation von mir selbst und ausschließlich mit den angegebenen Hilfen verfasst, keine anderen als die angegebenen Quellen benutzt und alle übernommen Zitate als solche gekennzeichnet wurden.

Diese Dissertation wurde in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Ort, Datum	Niklas Farke

II Abstract

Microbial metabolism is feedback regulated on many layers. Feedback control can be executed by metabolite-protein interactions to allosterically control enzyme activity, and to transcriptionally control enzyme abundance, providing cells with robustness to withstand, and adapt to perturbations. However, these interactions contribute to the complexity of metabolism and prohibit an intuitive understanding. To gain a deeper understanding of metabolism, mechanistic mathematical models are useful tools to simplify complex interrelationships. In this thesis, we developed mathematical models to study allosteric feedback, transcriptional feedback, and the interplay of both mechanisms. Then, we used this knowledge to develop a method to map feedback regulation between metabolism and transcription in *E. coli* metabolism. Since high-quality metabolite data are crucial to study metabolism, we finish this thesis with two chapters on mass spectrometry-based metabolomics.

After providing a general introduction in **Chapter 1**, we develop a mathematical model of amino acid biosynthesis in **Chapter 2** to study the interplay between allosteric feedback and transcriptional feedback. We showed that both feedbacks act in concert to balance robustness and efficiency. In **Chapter 3**, we developed a mathematical model of glycolysis that is transcriptionally regulated by the transcription factor Cra, to study metabolic burden in glycerol-producing *E. coli*. A robustness analysis revealed that Cra regulation causes growth defects and low glycerol titers in *E. coli*, and that this burden can be solved by engineering Cra regulation into the glycerol pathway. In **Chapter 4**, we developed a mathematical model to understand the implications of an ornithine-based allosteric activation at the branch point between arginine and pyrimidine biosynthesis. We showed that the feedback activation buffers upstream perturbations and thereby stabilizes pathway end products. In **Chapter 5** we investigated causes for periodic pyruvate oscillations using a mathematical model of glycolysis. We show that feed forward activation of the pyruvate kinase and high saturation of the pyruvate dehydrogenase contribute to pyruvate oscillations. In **Chapter 6** we performed knockdowns of 283 genes of *E. coli* metabolism and measured proteome and metabolome of the perturbed strains. A pathway-based analysis allowed us to map feedback regulation between metabolism and transcription using proteome and

metabolome data. In **Chapter 7** we analysed and validated the mass spectrometry-based flow-injection metabolomics approach with 160 spike-in samples. We showed that flow-injection causes complex MS1 spectra that can lead to false positive peak annotations. Finally, we concluded this thesis in **Chapter 8** by developing an approach to generate reference fragments for low-abundant, or commercially unavailable metabolites to complement reference databases. We showed a proof of principal for two metabolites.

III Zusammenfassung

Der mikrobielle Stoffwechsel ist komplex und wird durch Feedback Regulation auf vielen Ebenen reguliert. Feedback Regulation kann durch Metabolit-Protein Interaktionen ausgeführt werden, die allosterisch Enzymaktivität, oder transkriptionell Enzymmengen regulieren. Diese Regulationsmechanismen erlauben es den Zellen Störeinflüssen zu widerstehen und sich an die neuen Gegebenheiten anzupassen. Jedoch tragen diese Interaktionen zur Komplexität des Stoffwechsels bei, was ein Intuitives Verständnis erschwert. Um ein tieferes Verständnis des Stoffwechsels zu erreichen können mathematische Modelle genutzt werden, um komplexe Sachverhalte zu vereinfachen. In dieser Arbeit entwickeln wir mathematische Modelle um allosterische *Feedback* Regulation, transkriptionelle *Feedback* Regulation und das Zusammenspiel beider Mechanismen zu verstehen. Darüber hinaus nutzen wir dieses Wissen, um *Feedack* Mechanismen zwischen Metabolismus und Transkription zu kartieren. Da hochwertige Metabolit Daten für diese Aufgaben erforderlich sind, beschäftigen wir uns zum Schluss mit Massenspektrometrie-basierter Metabolomik.

Nach einer allgemeinen Einleitung in **Kapitel 1** entwickeln wir in **Kapitel 2** ein mathematisches Modell des Aminosäurestoffwechsels, um das Zusammenspiel zwischen allosterischer Regulation und transkriptioneller Regulation besser zu verstehen. Wir zeigen, dass beide Feedback Mechanismen benötigt werden um ein Gleichgewicht zwischen Robustheit und Effizienz einzustellen. In **Kapitel 3** entwickeln wir ein transkriptionell reguliertes mathematisches Modell der Glykolyse, um die metabolische Belastung von Glycerol produzierenden *E. coli* zu verstehen. Wir zeigen, dass transkriptionelle Regulation durch Cra in Glycerol produzierenden *E. coli* zu niedrigen Wachstumsraten und Glycerol Titern führt und dass konstruierte Cra Regulation der Glycerol Synthese die Wachstumsrate stabilisiert und Glycerol Titer erhöht. In **Kapitel 4** entwickeln wir ein mathematisches Model, um den Einfluss einer Ornithin-basierten allosterischen *Feedback* Aktivierung auf den Knotenpunkt der Arginin- und Pyrimidin Biosynthese zu charakterisieren. Wir zeigen, dass die *Feedback* Aktivierung Perturbationen stromaufwärts kompensiert und dadurch die Endprodukte der Arginine und Pyrimidin Biosynthese stabilisiert. Anschließend entwickeln wir in **Kapitel 5** ein mathematisches Modell der Glykolyse und untersuchen dieses auf die

Ursachen von Pyruvat Oszillationen. Wir zeigen, dass die *Feedforward* Regulation der Pyruvatkinase, sowie das Sättigungsverhalten der Pyruvat Dehydrogenase entscheidend sind für die Entstehung von Pyruvat Oszillationen.

In **Kapitel 6** regulieren wir 283 metabolische Gene im Stoffwechsel von *E. coli* genetisch herunter und messen anschließend Protein- und Metabolitänderungen. Mit Hilfe dieser Daten kartieren wir dann *Feedback* Regulation zwischen Metabolismus und Transkription in *E. Coli*.

In **Kapitel 7** analysieren und validieren wir die Massenspektrometrie-basierte Fließeinspritzung, indem wir *E. coli* Extrakte mit 160 individuellen Metabolit Standards versetzen. Wir zeigen, dass die Methode komplexe Spektren erzeugt, die zu falsch-positiven Peak Annotierungen führen können.

In **Kapitel 8** beenden wir diese Arbeit mit der Entwicklung einer Methode zur Generierung von Referenz Fragmenten für niedrig abundante, oder kommerziell nicht verfügbare Metabolite, um Referenzdatenbanken zu komplementieren. Wir zeigen das Prinzip der Methode für zwei Metabolite.

IV Acknowledgements

With the end of my Ph.D., a five-year journey ends in which I met incredible people in Marburg and Tübingen. I would like to express my gratitude to everyone for the great time.

First and foremost, I would like to thank **Hannes** for showing faith in me and letting me join your group as computational Ph.D. student. I could not be happier with the way you managed the group and me over this period. I value your personal and scientific advice and I have learned a lot under your guidance.

Next, I want to thank **Thorben**, **Paul,** and **Andi** for the great time in Marburg and Tübingen. It is rare for me to find friends that share the same hobbies and sense of humour. We can talk about private stuff, then go completely nuts over some random topic, and in the next minute we can have a critical scientific discussion. Thank you for making free time and work more enjoyable. Thank you, **Martin**, **Timur**, **Stefano**, **Dominik**, **Michelle**, **Chun-Ying** and **Dusica** for the great time in Marburg. It would not have been the same experience without you. I loved working with you and going out to the Oberstadt with you to enjoy a few drinks. I also want to thank **Michal** and **Marc** for being part of the best office on the second floor. Next, I want to thank **Alejandra, Vanessa, Sevvalli, Felicia, Johanna, Elisabeth**, **Michaela**, and all people from **Functional Metabolomics Lab** for the great time in Tübingen. Moving here during Corona was tough, but each one of you contributed to making it a positive experience.

I also want to thank **Georg Fritz**, **Jan Hasenauer, Lennart Randau**, and **Peter Lenz** for being part of my Thesis Advisory Committee, and for the scientific discussions and advises you provided.

Finally, I want to thank my parents for always supporting me during this journey.

V Table of Contents

I Eigenständigkeitserklärung ... ii

II Abstract ... iv

III Zusammenfassung ... vi

IV Acknowledgements ... viii

V Table of Contents ... x

Chapter 1 General introduction ... 14

1.1 Structure and function of the metabolic network of *Escherichia coli* ... 14
1.2 Systems biology of metabolism and mathematical modelling ... 22
1.3 Modelling metabolism and transcriptional regulation ... 29
1.4 Mass spectrometry-based metabolomics ... 32
1.5 Goal of this thesis ... 36
1.6 References ... 37

Chapter 2 A mathematical model predicts a trade-off between robustness and efficiency in amino acid biosynthesis ... 48

2.1 Abstract ... 49
2.2 Introduction ... 49
2.3 Allosteric feedback inhibition causes enzyme overabundance in amino acid biosynthesis ... 51
2.4 Interplay of allosteric feedback inhibition and enzyme-level regulation ... 53
2.5 Allosteric feedback and transcriptional feedback balance robustness and efficiency in amino acid biosynthesis ... 56
2.6 Oscillations in amino acid biosynthesis as by-product of robust efficiency ... 58
2.7 Enzyme overabundance provides robustness against genetic perturbations ... 60
2.8 Discussion ... 61
2.9 Methods ... 63
2.10 References ... 66

Chapter 3 A mathematical model predicts robust regulation in engineered glycerol-producing *E. coli* ... 70

3.1 Abstract ... 71
3.2 Introduction ... 71
3.3 Glycerol overproduction in *E. coli* causes growth burden ... 73
3.4 A mechanistic model predicts robust transcriptional regulation for glycerol producing *E. coli* ... 75
3.5 A Cra-regulated pBAD promoter improves growth rate and glycerol titers ... 79
3.6 Discussion ... 80
3.7 Methods ... 82
3.8 References ... 86

Chapter 4 A mathematical model reveals the function of an ornithine-based allosteric activation of CarAB ... 90

4.1 Abstract ... 91
4.2 Introduction ... 91
4.3 CRISPRi knockdown of CarAB causes ornithine accumulation ... 92

4.4 Feedback activation of CarAB by ornithine stabilizes pathway end-products in arginine and pyrimidine biosynthesis 93
4.5 Discussion 97
4.6 Methods 98
4.7 References 100

Chapter 5 A mathematical model of glycolysis predicts sustained pyruvate oscillations 104

5.1 Abstract 105
5.2 Introduction 105
5.3 A mathematical model of *E: coli* glycolysis predicts sustained pyruvate oscillations 106
5.4 Discussion 110
5.5 Methods 111
5.6 References 113

Chapter 6 Mapping regulation in *E. coli* metabolism with a metabolism wide CRISPRi library 116

6.1 Abstract 117
6.2 Introduction 117
6.3 CRISPRi knockdowns cause downregulation of target genes 118
6.4 Proteome changes reveal feedback loops in 24 metabolic pathways 120
6.5 Transcriptional regulation in proximity of the effector metabolite 124
6.6 CRISPRi knockdowns cause substrate accumulations 126
6.7 Screening for metabolites that control gene expression using untargeted metabolomics 128
6.8 Discussion 131
6.9 References 132

Chapter 7 A network approach identifies in-source modifications of primary metabolites during flow-injection mass spectrometry 136

7.1 Abstract 137
7.2 Introduction 138
7.3 FI-MS with 160 authentic metabolite standards 140
7.4 Single metabolites can produce extensive in-source derivates 142
7.5 A network approach explains significant *m/z* features in FI-MS spectra 145
7.6 MS2 information identifies significant features that are in-source fragments 147
7.7 Misannotation of in-source derivates to metabolites 149
7.8 Discussion 150
7.9 Material and Methods 151
7.10 References 153

Chapter 8 Acquisition of unknown MS2 fragments for low abundant and unavailable metabolites with CRISPR interference 158

8.1 Abstract 159
8.2 Introduction 159
8.3 FI-MS identifies significant metabolome changes of pathway intermediates in purine, histidine and isoprenoid biosynthesis 160
8.4 Identification of MS2 fragments for metabolites in histidine, purine, and isoprenoid metabolism with an LC-MS/MS approach 163
8.5 Discussion 165
8.6 Methods 166
8.7 References 167

Chapter 9 Key findings **170**
Chapter 10 Conclusions and Outlook **174**
10.1 References 176
VI Curriculum Vitae ..... **clxxviii**
VII Publications of this thesis ..... **clxxix**
VIII Abgrenzung der Eigenleistung ..... **clxxx**
IX List of tables ..... **clxxxii**
X List of figures ..... **clxxxiii**
XI List of Abbreviations ..... **clxxxv**
XII Appendix ..... **clxxxvi**

Chapter 1 General introduction

1.1 Structure and function of the metabolic network of *Escherichia coli*

1.1.1 Structure and function of the metabolic network

Metabolism is executed by a network of biochemical reactions that convert nutrients via various intermediary metabolites into energy and building blocks that are required for cell growth. Despite 3.7 billion years of divergent evolution, the basic structure of metabolic networks is highly conserved across the tree of life, and basically all life forms rely on a similar set of metabolites[1]. Metabolites allow the cell to sense environmental changes, provide a means for cell-to-cell communication, and can act as defense mechanism[2]. For all living organism, the structure of the metabolic network is encoded on their genomes. According to the central dogma of molecular biology, genes are transcribed into mRNA, which are then translated into enzymes that catalyzes chemical reactions with high selectivity[3]. Together, multiple enzymatic reactions are organized in modular, yet densely connected metabolic pathways[2]. The pathways can be broadly grouped into catabolism and anabolism[4,5] (**Figure 1**). Catabolism breaks down nutrients to produce anabolic building blocks and energy in the form of adenosine triphosphate (ATP). For example, glycolysis is an important catabolic pathway that converts glucose to pyruvate and energy. Anabolism uses intermediates and energy to produce building blocks required for cell growth. An example for anabolism is amino acid biosynthesis. For each of the 20 proteinogenic amino acids, precursors from central metabolism are converted into the final amino acids by multiple enzyme-catalyzed reaction steps.

For the well-studied model organism *Escherichia coli* (*E. coli*) the structure and stoichiometry of the metabolic network is assumed to be largely complete[6,7]. Despite that, we know little about how *E. coli* maintains and adapts its metabolism dynamically in changing environments. Filling this knowledge gap is important to design better production strains in biotechnology[8], to help in understanding microbial communities[9], and to help in understanding diseases with metabolic phenotypes[10].

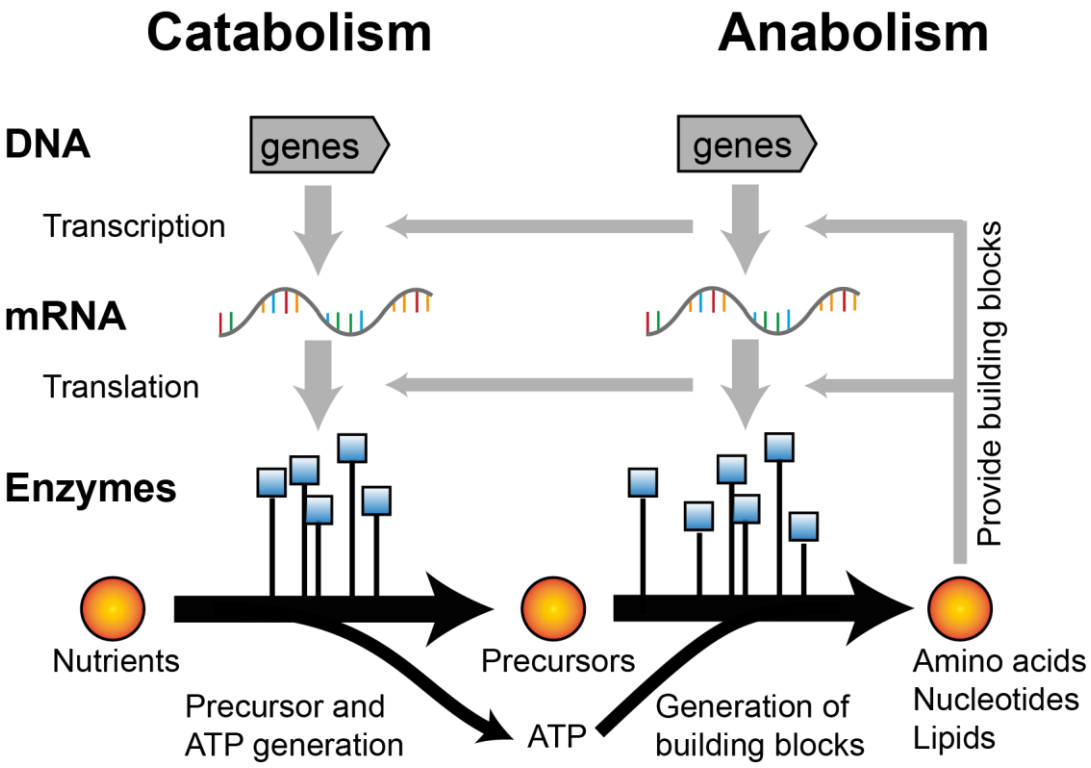


Figure 1. Central dogmatic flow of information, catabolism, and anabolism*. Genes are pointed boxes. Enzymes are blue squares. Metabolites are orange spheres. Black arrows are metabolic reactions. Grey arrows indicate the information flow as defined in the central dogma of molecular biology. Genes are transcribed into mRNA. mRNA is translated into enzymes. Enzymes catalyse biochemical reactions. In catabolism, the enzymes convert nutrients into ATP and precursor metabolites. In anabolism, enzymes produce biosynthetic precursors (amino acids, nucleotides, lipids) from precursors and ATP. The biosynthetic precursors are required for transcription and translation.*

1.1.2 Robustness of metabolism

Cells constantly encounter internal and external perturbations, such as temperature gradients and changes of nutrient availability. To survive, grow, and evolve, they need to be robust enough to cope with these perturbations. Robustness is an emergent property of biological systems that allows them to maintain function amidst various perturbations[11,12].

To understand metabolism, it is crucial to understand the mechanisms that ensure robustness. Although there is no unifying way to describe robustness theoretically or experimentally yet[13], there are simple mechanisms that enable metabolic robustness. These general mechanisms are based on **system control**, **modularity**, and **redundancy**[11,12] (**Figure 2a**). **System control** uses negative and positive metabolic feedback to adjust metabolic rates and metabolite levels to new conditions.

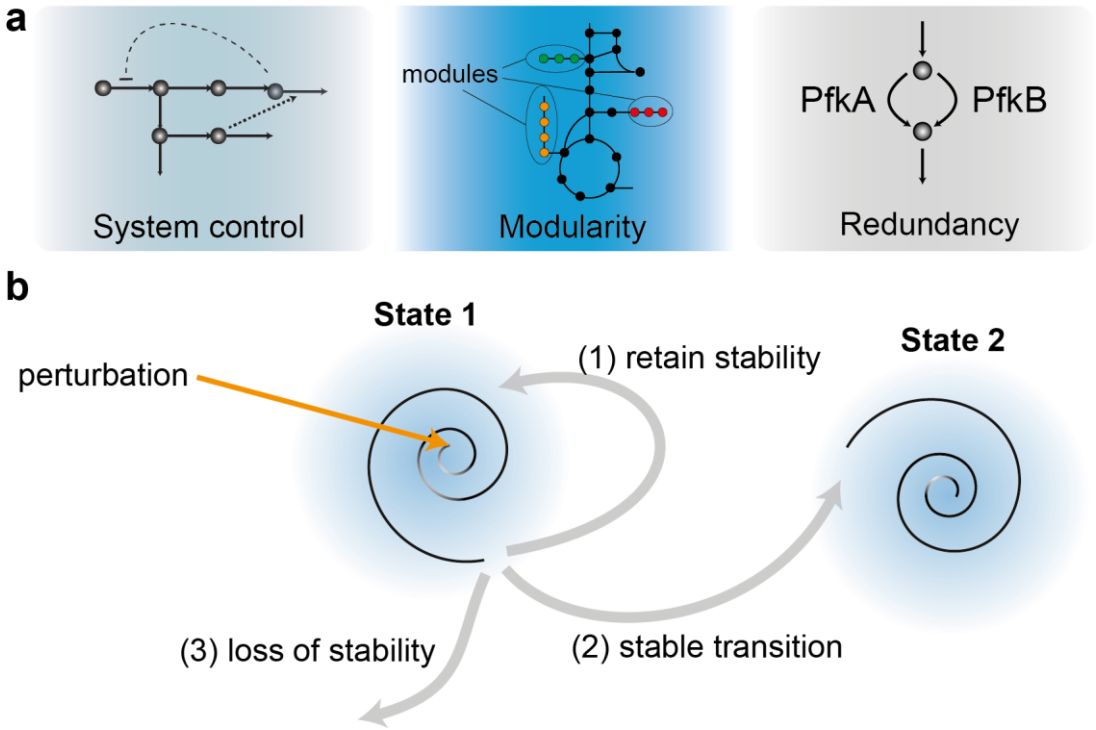


Figure 2. Principles of biological robustness. a) *Examples of system control, modularity, and redundancy in metabolic networks. Spheres are metabolites and solid arrows are metabolic reactions. (Left) System control is achieved by negative feedback (blunt dashed arrow) and positive feedback (pointed dashed arrow). (Middle) Modularity is achieved by spatial and functional separation of metabolic pathways. Different pathway modules are coloured. (Right) Redundancy is achieved by different enzymes that catalyse the same reaction, for example PfkA and PfkB in glycolysis.* ***b)*** *Example of how a perturbation affects a cellular state. (1) The system returns immediately to the same steady state. (2) The system loses stability transiently (State 1), and transitions to a new steady state (State 2). (3) the system loses stability and becomes unstable.*

One common way metabolism executes **feedback control** is via metabolite-protein interactions, in which binding of a metabolite modulates protein activity[7]. While negative feedback enables robust adaptation[14], positive feedback enhances sensitivity, allowing the system to switch between different states[11].

Redundancy is based on alternative paths that can carry out the same function. For example, ATP can be produced via oxidative phosphorylation and by glycolysis. While oxidative phosphorylation requires oxygen, glycolysis can still produce ATP under anaerobic conditions, albeit less efficiently[12]. Another way to achieve metabolic redundancy is by gene duplications. For example, the genes *pfkA* and *pfkB* produce two isoforms of the phosphofructokinase (PFK) that have overlapping functions and catalyze the same reaction in glycolysis[15].

Modularity means that metabolism is composed of a diverse set of functional units. Due to functional and spatial separation, perturbations are kept local, which prevents system failure and cell death[16]. Moreover, modularity enables evolution by creating a robust core that is then augmented by additional modules, which often share a set of recurring network motifs[17,18].

Although these mechanisms are crucial for surviving, the cell faces hard trade-offs to achieve robustness[19]. Complex systems evolved robustness to handle certain conditions or perturbations. However, the same mechanisms that allow the systems to be robust against known perturbations render them fragile against unknown or rare perturbations, as defined by the theory of highly optimized tolerance[19–21].

For example, feedback regulation enhances robustness of a system in a common environment. However, when faced with unknown perturbations, the same feedback mechanisms can induce fragilities in the form of sustained oscillations, causing system failure[22]. Metabolic redundancy provides robustness via fail safe mechanisms, but maintaining redundancy requires resources. While redundancy protects the cell in certain environments, it also causes a fitness disadvantage in environments in which that specific type of redundance is not needed[23].

Robustness needs to be clearly distinguished from stability. While stability defines a certain state the system operates in, robustness is defined by functionality, irrespective of the state of the system. When a stable system is perturbed, there are basically three possible outcomes. First, the system returns to the original state. Second, the system loses stability, but transitions towards a new stable steady state. Third, the system becomes unstable. The first and second cases are examples of robust behavior, even though stability is transiently lost in the second case[12,13] (**Figure 2b**).

In conclusion, biological systems are robust, and this robustness is achieved by feedback control, modularity, and redundancy[11,12 22]. However, the more a system is optimized to withstand known perturbations, the more fragile it becomes against other perturbations[20]. Thus, understanding metabolic robustness and its trade-offs is key to understanding how cells adapt dynamically to changing environments, or perturbations in general.

1.1.3 Dynamic control of metabolism

One of the most prevalent mechanisms for cells to achieve robustness is feedback control[11,12]. Feedback control is how a cell manages all steps between gene expression and metabolism. Most commonly, feedback is executed via metabolite-protein interactions[7]. Briefly, metabolites can transiently bind to certain proteins. If a perturbation changes the concentration of that metabolite, saturation of the target protein changes, often inducing structural change that increases or decreases its activity[24].

Biochemical reactions are carried out by enzymes, and their reaction rate depends on enzyme specific parameters like the turnover number, and binding constants, and enzyme abundance. The rate at which one metabolite is converted to another metabolite defines the metabolic flux of a metabolic reaction[25]. It is generally assumed that metabolic networks of exponentially growing cells operate in a steady state, in which the concentration of each metabolite is time invariant[26]. For example, in linear metabolic pathways subsequent reactions must share the same metabolic flux. Perturbations, however, can cause a loss of a steady state, leading to metabolite accumulations or depletions, and a change in metabolic flux[27] (**Figure 2**). Since metabolic flux depends on enzyme abundance and enzyme activity, the cell can regulate flux by using feedback control to change enzyme activity, or enzyme abundance, or both[28].

One of the most important mechanisms to control enzymatic activity by metabolites is allosteric feedback regulation[7]. A different metabolite than the substrate or the product of the enzyme binds to a binding pocket of the same enzyme that is not the catalytic center. This metabolite-enzyme interaction induces a conformational change of the enzyme structure, affecting enzyme activity within seconds of the interaction[24]. Thus, allosteric feedback regulation allows the cell to modulate metabolic flux by changing enzyme activity in response to changes in metabolite levels (**Figure 3a**). Besides allosteric regulation, the activity of many enzymes is inhibited by metabolites that compete with the substrate for a place in the active site. These metabolites often structurally resemble the substrates[29]. Other mechanisms to control protein activity include acetylation, phosphorylation, and methylation. In contrast to transient

metabolite-protein interactions, these mechanisms can be either reversible or irreversible[30].

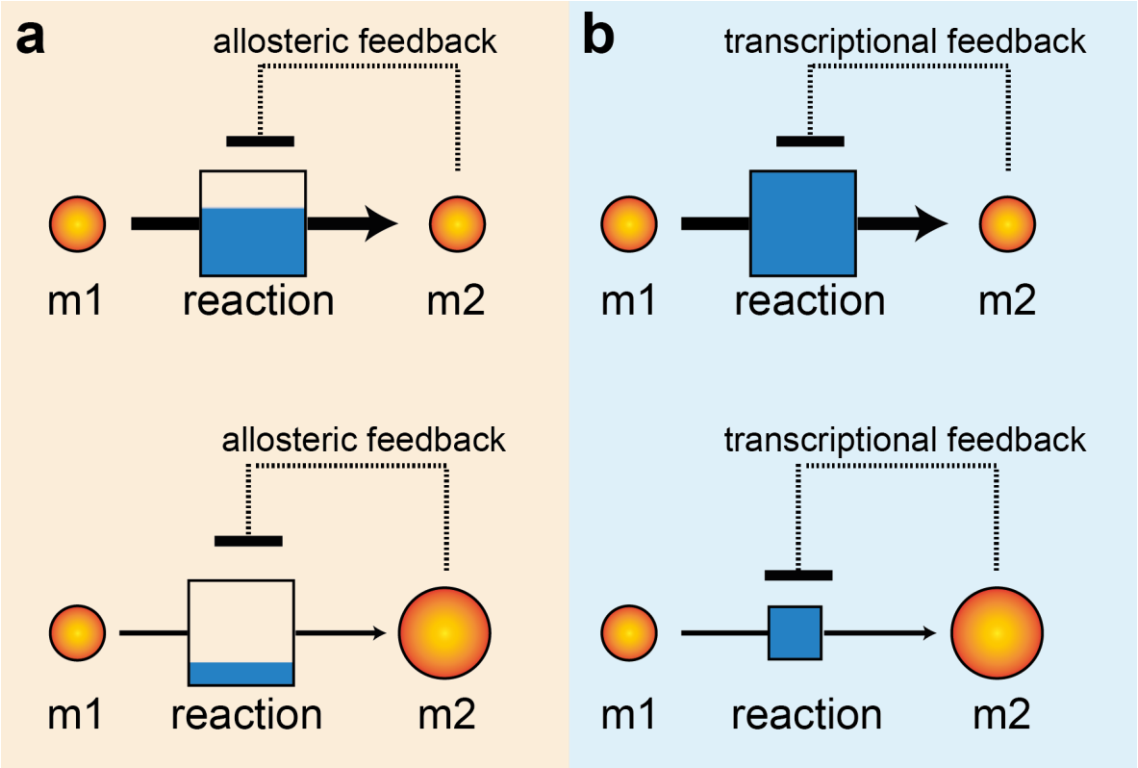


Figure 3. Allosteric feedback and transcriptional feedback*. Orange spheres are metabolites (m1 is the substrate, m2 is the product). Solid arrows are metabolic reactions. Squares are enzymes that catalyse the reactions. Dashed arrows are allosteric or transcriptional feedback regulations. Square and Sphere size corresponds to the enzyme amount and metabolite amount. The amount of fill corresponds to the activity of the enzyme. The thickness of the arrow corresponds to the magnitude of the reaction flux* ***a)*** *Allosteric feedback inhibition. An increase of m2 causes a decrease of the enzyme activity and a decrease of the reaction flux.* ***b)*** *Transcriptional feedback inhibition. An increase of m2 causes a decrease of the enzyme amount, and a decrease of the reaction flux.*

The metabolic flux of an enzyme-catalyzed reaction can also be modulated by changing enzyme abundance (**Figure 3b**). Enzyme abundance can be regulated transcriptionally by transcription factors (TFs)[31], transcriptional attenuation[32], or riboswitches[33], but also by changing the number of active ribosomes[34]. Transcription factors are organized hierarchically in the transcription regulation network (TRN)[35]. In the TRN there are local transcription factors, which control enzyme abundance of specific metabolic modules, and there are global transcription factors that control genes of many different modules[36]. Controlling enzyme abundance allows the cell to control its resources, and similarly to allosteric regulation, some transcription factors can sense changes in metabolite concentrations[36,37]. Binding of a metabolite modulates the activity of the TF

and thereby gene expression is controlled. For example, it was shown that the transcription factor Cra can sense glycolytic flux by interacting with the metabolite fructose-6-phosphate[37]. Other examples can be found in amino acid metabolism, in which the amino acid of a pathway controls enzyme levels of that pathway by interacting with a local transcription factor[36]. Besides activation of transcription factors by metabolite binding, transcription factors can also sense a change of the pH, or oxidative stress[38].

Thus, to quickly adapt to new internal or external cues, the cell can adjust metabolic flux by means of allosteric regulation,[24] and enzyme abundance by means of transcriptional regulation[39]. Both mechanisms have in common that metabolites are the key signaling molecules. Although allosteric and transcriptional regulation have been studied for a long time, many aspects of the dynamic control of metabolism remain elusive[7]. Especially the interplay between both mechanisms has not been thoroughly investigated yet[28].

1.1.4 Identification of metabolite-protein interactions

The interaction between metabolites and proteins causes a change in the physicochemical properties of the interactors, and approaches to try to identify interactions are usually based on changes in these properties. These approaches focus either on the protein or the metabolite[40].

Historically, *in vitro* binding assays, combining a single enzyme with a single metabolite, were carried out to identify binding constants from changes in protein properties[41,42]. Mass spectrometry-based proteomics and metabolomics approaches have been developed to study metabolite-protein interactions systematically at a higher throughput. For example, mass spectrometry-based proteomics was used to systematically identify structural protein changes upon metabolite bindings in cell lysates[43]. More recently, an approach based on limited proteolysis was used to identify a hitherto unprecedented number of metabolite-protein interactions in *E. coli*[44].

However, while many metabolite-protein interactions were discovered this way, the data are based on cell lysates, which may not be relevant *in vivo*. In contrast, mass spectrometry-based proteomics was used to measure temperature-based changes of protein stability in intact cells. This approach enabled the detection of multiple

metabolite-protein interactions that caused a change in protein stability[45,46]. Other approaches focus more on metabolite changes upon protein binding. For example, NMR was used to identify metabolites that interact with single proteins[47].

While these approaches identified metabolite-protein interactions, an interaction alone does not imply an underlying function. Functional interactions cause physiological changes in the cell. The function of an interaction can then be inferred by analyzing multi-omics data. A recent approach used a comprehensive multi-omics data set comprising 25 growth conditions in yeast and prior knowledge in form of simple mathematical models to identify functional metabolite-protein interactions[48].

To understand regulation, cells are perturbed, and the function of the regulation is then inferred from the differences between the original and the perturbed state. Common perturbations are changes of carbon sources[49], abiotic stresses, but also gene knockouts[50]. Systematic analyses to assign function to metabolite-protein interactions use multi-omics data. For example, correlating promoter activity with metabolomics data showed that only a few metabolites regulate up to 90 % of experimentally measured transcription changes[51]. In another study, allosteric regulators driving the transition between different carbon sources were identified from correlating flux, metabolome, and transcriptome data[49].

In conclusion, there are many approaches that allow for the identification of metabolite-protein interactions. However, to determine their relevance *in vivo*, information from multiple omics sources is required, and sometimes even prior knowledge in the form of kinetic enzyme parameters is needed[48]. Moreover, knockouts and environmental perturbations cause very strong, and often global cellular responses, which may prohibit the identification of subtle regulations[49]. Thus, to find subtle regulations, knockdowns may be more suitable perturbations. This could be achieved by CRISPR interference, which blocks transcription of a target gene, causing its dilution by growth[52]. Finally, mathematical models provide additional evidence that can be used in conjunction with data, or independently, to identify and study functional metabolite-protein interactions.

1.2 Systems biology of metabolism and mathematical modelling

1.2.1 Systems biology

Years of progress in molecular biology have greatly increased the coverage of all layers of cellular organization. Moreover, technological advances in genome editing[53], sequencing[54], and mass spectrometry[55] have enabled the research of complex biological interactions that surpass the function of individual components. To achieve a system-level understanding, it is crucial to understand the design principles of biological networks, and how these networks are controlled dynamically[56]. However, even though it is now possible to measure system-level data, the non-linear relationships between different components prohibits an intuitive understanding and poses an obstacle during data interpretation.

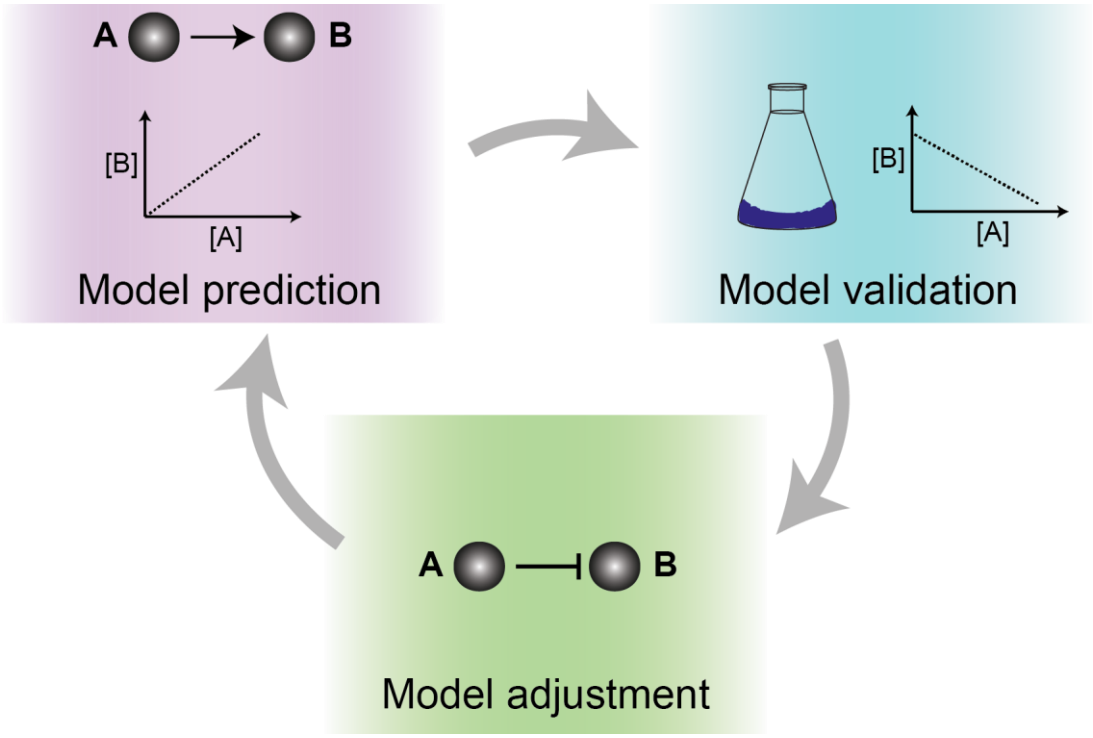


Figure 4. Model-experimentation cycle*. In the beginning there is a conceptual hypothesis or a model prediction about a process (purple). This hypothesis can be validated experimentally (blue). If the experiment shows a different behaviour, the model needs to be adjusted (green). Then predictions are made with the new model structure.*

To understand complex interactions, and to organize disparate information, mathematical models are invaluable tools[57]. Mathematical models are reduced, or abstracted simplifications of the real systems, and focus on the essential components

and interactions. As such, they allow us to contextualize data, to provide a means to test, to verify, and to generate new hypotheses (**Figure 4**)[58]. They allow insights into emergent system properties that are difficult to assess experimentally, like stability[14] and robustness[13]. Especially in the field of metabolism, mathematical models already had a huge impact on our system understanding. These models can take on various complexities, ranging from linear models to coarse grained, or mechanistic kinetic models[4,27,59–61].

1.2.2 From stoichiometric to mechanistic models

Stoichiometric models. A simple way to model metabolic networks are stoichiometric models. These models employ the metabolite-reaction stoichiometry that is based on genome-scale metabolic reconstructions[6,62]. The latest metabolic reconstruction of *E. coli* contains 1,515 metabolic genes, 1,192 unique metabolites and 2,719 metabolic reactions[6]. A modelling approach that employs network stoichiometry to calculate metabolic fluxes is Flux Balance Analysis (FBA)[61]. FBA assumes a steady state in which all metabolic concentrations and reaction fluxes are time invariant. Linear programming is then used to find a flux distribution that optimizes a biologically relevant objective function (e.g. maximizing the growth rate)[63]. FBA has been successfully used to quantitatively predict growth rates[64], gene essentiality,[65] and production rates/yields of commercially valuable metabolites[66]. Although FBA predicts metabolic fluxes with high accuracy, it does not allow to study system dynamics. Therefore, FBA is unsuitable to answer how fluxes are achieved[60], nor is it suited to study dynamic control of metabolism.

To study the dynamics of metabolism, kinetic models based on ordinary differential equations (ODEs) can be used. There are coarse-grained kinetic models that describe dynamic systems phenomenologically, and there are mechanistic kinetic models that try to describe the molecular interactions in detail[60]. Recently, coarse grained models have described fundamental principles of bacterial physiology[4,67,67,68]. For example, a simple model with feedback inhibition and feed-forward activation predicted that keto acid abundance controls the cAMP-dependent catabolite repression in *E. coli*[4]. Although coarse-grained models provide a basic understanding of metabolism, they are unsuitable to study detailed molecular mechanisms[60].

Mechanistic models. Mechanistic models are the most detailed mathematical representations of metabolism. For example, mechanistic models of metabolic pathways employ network stoichiometry, and detailed kinetic rate-laws for each enzymatic reaction step. As such, they allow to study dynamic interactions like allosteric regulation[69–71]. Over the years, rate laws of various complexities have emerged to model enzymatic reaction rates[69,70,72].

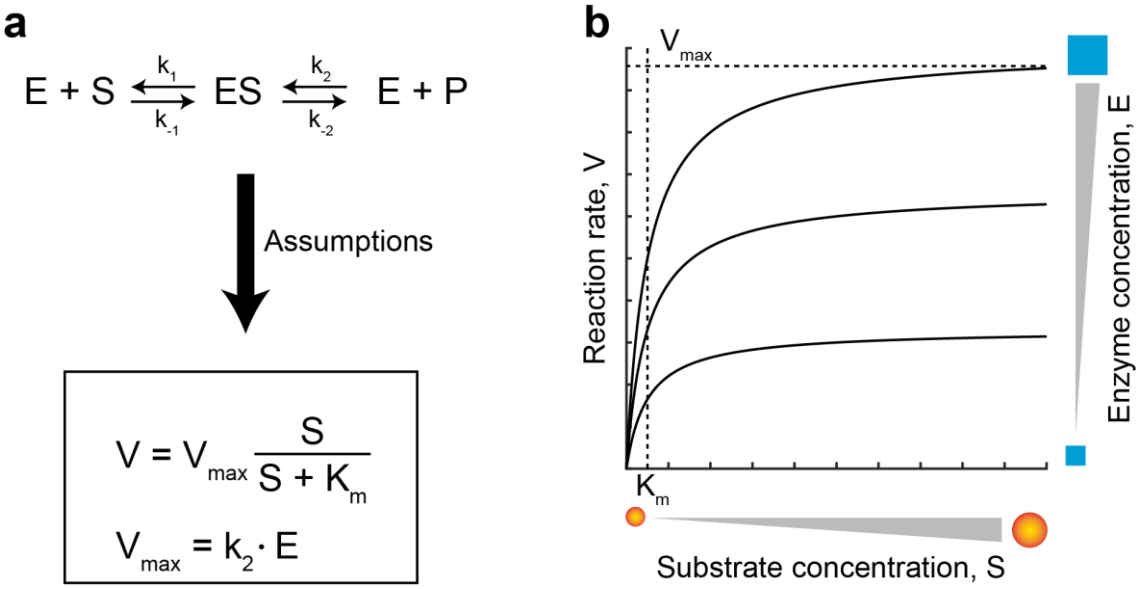


Figure 5. Michaelis-Menten equation. a) *Reaction scheme of the Michaelis-Menten equation. S is the substrate of the enzyme E. ES is the enzyme substrate complex. P is the product of the reaction. k are mass action reaction constants. Assumptions lead to the simplified irreversible Michaelis-Menten equation with the parameters Vmax and Km. Vmax is the product of the turnover number k2 and the enzyme amount E.* ***b)*** *Michaelis-Menten plot. The reaction rate V is plotted against the substrate concentration S. The curve asymptotically approaches Vmax. At Vmax/2 the substrate concentration equals Km.*

One of the best-known mechanistic rate laws is the Michaelis-Menten equation (**Figure 5a**), which describes a hyperbolic relationship between substrate concentration and reaction rate (**Figure 5b**)[73]. The conversion of the substrate requires several elemental steps, which are modelled as a system of ODEs using mass-action kinetics[72]. To allow for a tractable model analysis, the model is simplified by making assumptions[74]. The simplifications assume a steady state, excess of substrate and an irreversible product formation rate. These assumptions yield a simplified kinetic model with two lumped kinetic parameters (K_m, V_{max}), which can be determined *in vitro (***Figure 5*a*)*. Because of its simplicity and practical relevance, Michaelis-Menten kinetics have been used frequently to model metabolic reactions. For example, the kinetic model k-ecoli457 was

modelled using mainly Michaelis-Menten kinetics[75]. Here, the predicted Michaelis-Menten constants matched experimentally determined parameters[75]. However, for many enzymes the rate laws and kinetic parameters remain unknown and *in vitro* data may not be appropriate to study *in vivo* systems[76]. This lack of knowledge makes it difficult to build accurate mathematical models.

1.2.3 Addressing model uncertainties with ensemble modelling

Mathematical models of metabolism suffer from epistemic uncertainty, that is the lack of knowledge about interactions, rate-laws, and kinetic parameters[77]. Even though, recent studies suggested that kinetic parameters may be less important than model structure, these uncertainties still prohibit the design of accurate mechanistic models[78,79].

Ensemble modelling has been used to identify model parameters, and to find new physical interactions. The common denominator of all ensemble modelling approaches is the formulation of multiple alternative models that represent different hypotheses[71,80,81]. The idea is that an ensemble of multiple models is more accurate than only a single model. For example, the models of an ensemble can all differ in their kinetic parameters (**Figure 6a**), or they can differ in their physical interactions (**Figure 6b**).

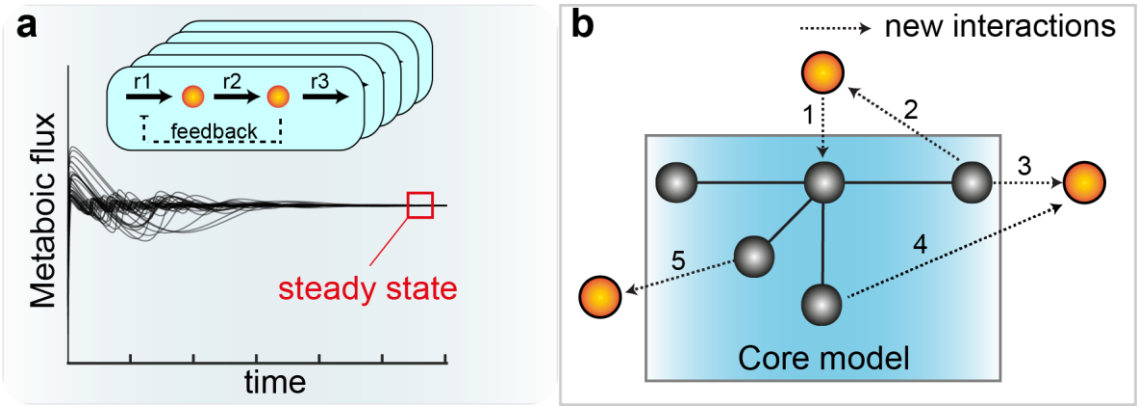


Figure 6. Ensemble modelling. a) *Ensemble of a kinetic metabolic model with two metabolites (orange spheres), three reactions (solid arrows) and allosteric feedback (dashed arrow). Blue boxes are models of the ensemble with the same model structure but different kinetic parameters. The graph shows the metabolic flux time course of the ensemble. All models reach the same steady state flux, but they have different time courses.* ***b)*** *Ensemble of structural models. Grey spheres are known components of the system and solid lines are known interactions between the known components (core model). Orange spheres are components for which there are no known interactions. Enumerated dashed arrows are potential interactions. Each model in the ensemble contains the core model and one additional new interaction.*

To identify kinetic parameters of a mechanistic model, a metabolic ensemble modelling approach constructs several models (an ensemble of models) that share the same structure and stoichiometry. Then, all models of the ensemble are constrained to the same steady state flux[27,71]. Free parameters are then calculated by random sampling from biologically relevant intervals, and the remaining parameters are calculated. Thus, all models in the ensemble reach the same steady state flux, but because each model has a unique set of kinetic parameters, the respective dynamics are different (**Figure 6a**). By constraining all fluxes to the same steady state flux, the number of feasible kinetic parameters is reduced to biologically relevant parameters[71]. In the original study, the models were then tested against experimental data by simulating gene expression changes. Only 1/1010 parameter correctly predicted the experimental data, and the single model was then selected to make further predictions[71]. In a different application, metabolic ensemble modelling has been used to study robustness of (engineered) metabolic networks against strong changes in enzyme concentrations.[82]

Besides knowledge-gaps in kinetic parameters and rate laws, basic physical interactions may be unknown. To identify unknown interactions in the mTOR pathway of *Saccharomyces cerevisiae (S. cerevisiae)*, a core model based on current knowledge was built. The core model was then extended by single additional interactions (**Figure 6b**). That resulted in an ensemble of 18 models that shared the same model core and differed only in a single interaction. Comparing the model simulations to experimental data then revealed the interaction that best explained the data[81].

Moreover, ensemble modelling was used to identify functional allosteric regulations. In this study, glycolysis was modelled with an ensemble of 126 different combinations of allosteric feedback regulations. Together with experimental data, the approach revealed that pyruvate allosterically activates the fructose-1,6-bisphosphatase, and that this interaction is crucial for *E. coli* to reversibly switch between gluconeogenesis and glycolysis[83].

In conclusion, mechanistic models of biological processes suffer from uncertainties in model parameters and in model structures. These uncertainties can be addressed by ensemble modelling techniques. Ensemble modelling uses multiple models with different kinetic parameters, or different model structures to make predictions, which

acknowledges uncertainty, and allows for more general predictions than using only a single model[71,81,84].

1.2.4 Stability and robustness of dynamic systems

As already discussed in **Chapter 1.1.2**, stability and robustness are emergent properties of biological systems that are crucial for the cell to survive and adapt[11,12]. However, while stability and robustness are difficult to assess experimentally, mathematics provide the means to define and analyse these emergent system properties[57].

Stability. The linear stability of a system of ODEs is defined at a fixed point or steady state, in which all relevant system variables are time invariant. Then, a variable of the system is perturbed by a small amount. If the system returns to the same state, the system is linear stable, otherwise the system is linear unstable. This behaviour is illustrated in **Figure 7a**. The function has two steady states (blue and orange sphere). The arrows indicate the velocity (dX/dt) of X when the steady state is perturbed slightly. The blue sphere always returns to the steady state (stable) and the orange sphere moves away from the steady state (unstable).

Mathematically, stability can be assessed by approximating the system at the fixed point by a Taylor series. The Taylor series is then cut after the linear term, and the eigenvalue of that term defines if the system is linear stable or unstable. For ODEs of order two or higher, the Jacobian matrix, a system-wide linearization, is calculated. Consider a system with two coupled equations (f1, f2), two variables (x_1, x_2), and kinetic parameters (p) that is in steady state:

$$\frac{dx}{dt} = F\left(x_{1_{SS}}, x_{2_{SS}}, p\right) = \begin{matrix} f_1(x_{1,SS}, x_{2,SS}) \\ f_2(x_{1,SS}, x_{2,SS}) \end{matrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{1.1}$$

The system is linearized around the steady state by calculating the partial derivatives with respect to the variables x1 and x2. This is the Jacobian matrix, J:

$$J = \frac{\partial F(x_{SS}, p)}{\delta x} = \begin{bmatrix} \frac{\delta f_1}{\delta x_1} & \frac{\delta f_1}{\delta x_2} \\ \frac{\delta f_2}{\delta x_1} & \frac{\delta f_2}{\delta x_2} \end{bmatrix} \tag{1.2}$$

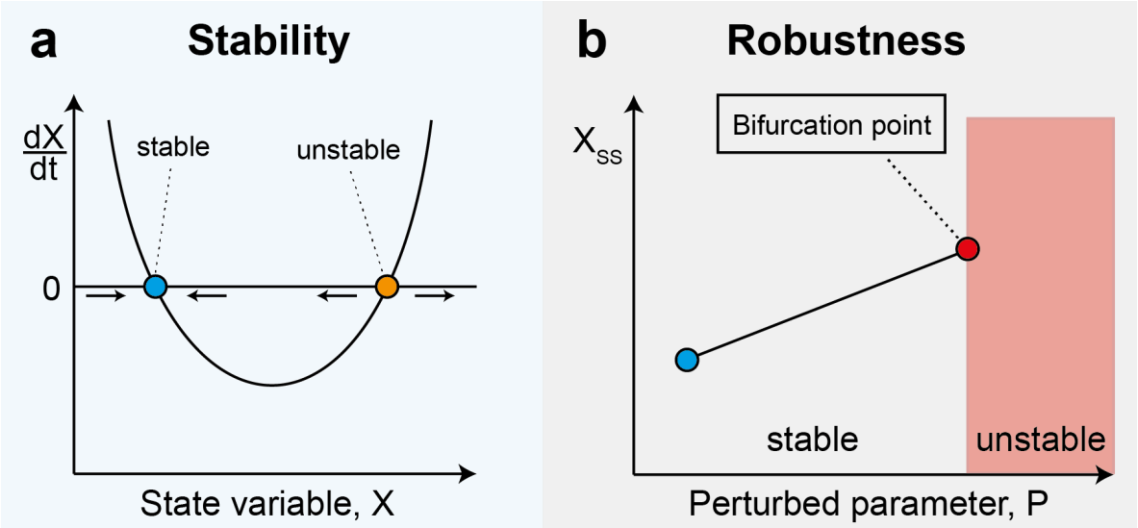


***Figure 7. Stability and Robustness**. **a)** Graphical description of linear stability. The first temporal derivative of X (dX/dt) is plotted against X. The function has two steady states (dX/dt = 0, blue and orange sphere). The arrows denote the direction the sphere moves when X is perturbed slightly. The blue steady state is stable, because the sphere moves back to the steady state after it is perturbed. The orange steady state is unstable, because the orange sphere moves away from the steady state after a perturbation. **b)** Bifurcation analysis. X_{SS} is the state variable X in steady state. The blue sphere is the starting steady state of the analysis. Each point of the black line contains stable steady states for different values of the parameter P. The red sphere indicates a bifurcation point at which the system becomes unstable (red area).*

The linear stability of the system is then determined by the eigenvalues of the Jacobian matrix. The system is linear stable if all eigenvalues are negative. The system is linear unstable, if at least one eigenvalue is positive[85].

Robustness. Robustness is a property that preserves functionality despite perturbations[11,12]. In mathematical models, perturbations can be simulated by varying a kinetic parameter. In a bifurcation analysis, the system is initially in a stable steady state. Then, as a parameter p is varied, the system moves along a plane of stable steady states. The system becomes unstable at the bifurcation point that defines the boundary between a stable and unstable parameter space (**Figure 7b)**. Robustness can then be defined as the degree of the perturbation that can be tolerated until the system becomes unstable[82]. For example, consider an ODE in steady state, in which x_{SS} are the variables in steady state and p is a kinetic parameter that is varied iteratively.

$$\frac{dx}{dt} = F(x_{SS}, p) = 0 \tag{1.3}$$

In steady state, the derivative of $F(x_{SS},p)$ with respect to the parameters is also zero:

$$\frac{dF(x_{SS},p)}{dp} = \frac{\delta F}{\delta x_{SS}} \cdot \frac{dx_{SS}}{dp} + \frac{\delta F}{\delta p} = 0 \quad (1.4)$$

Rearranging the equation yields an expression that describes changes in a steady state as a parameter p is varied iteratively.

$$\frac{dx_{SS}}{dp} = -\left(\frac{\delta F}{\delta x_{SS}}\right)^{-1} \cdot \frac{\delta F}{\delta p} \quad (1.5)$$

This expression contains the inverse of the Jacobian Matrix. If the Jacobian matrix has full rank, matrix inversion is possible. A bifurcation point is reached when the Jacobian matrix becomes singular. At this point, matrix inversion is no longer possible, and the system becomes unstable. In conclusion, stability and robustness of mathematical models can be analysed by linear stability analysis and bifurcation analysis[82].

Linearization of a model at a steady state resembles weak perturbations, and parameter continuation by bifurcation analysis resembles strong perturbations. These concepts have been used to study dynamic systems. For example, Metabolic Control Analysis (MCA) studies the global consequences of small local perturbations[86]. Additionally, parameter continuation methods have been used to simulate gene overexpression in the context of metabolic engineering[82].

1.3 Modelling metabolism and transcriptional regulation

Metabolic networks consist of multiple enzyme-catalysed reactions that all have a specific metabolic flux that is balanced throughout the network during exponential growth[61]. The metabolic flux of a metabolic reaction depends on specific kinetic enzyme parameters (e.g. Km, and the turnover number) and enzyme abundance[70,73,74]. When metabolism is perturbed by internal or external cues, the flux often needs to be adjusted by the cell to avoid accumulation or depletion of metabolites[82]. The flux of single enzymatic reactions can then be modulated by allosteric feedback control, and by changing enzyme abundance via transcriptional regulation[7] (**Figure 8**). However, although both mechanisms are mediated by metabolite-protein interactions, it is still unclear how they work together[28]. For example, it was shown that changes of transcriptional rates are bad predictors of metabolic flux[87], and the reason for this

behaviour is likely that enzymes are overabundant and usually not saturated by their substrates[88–90].

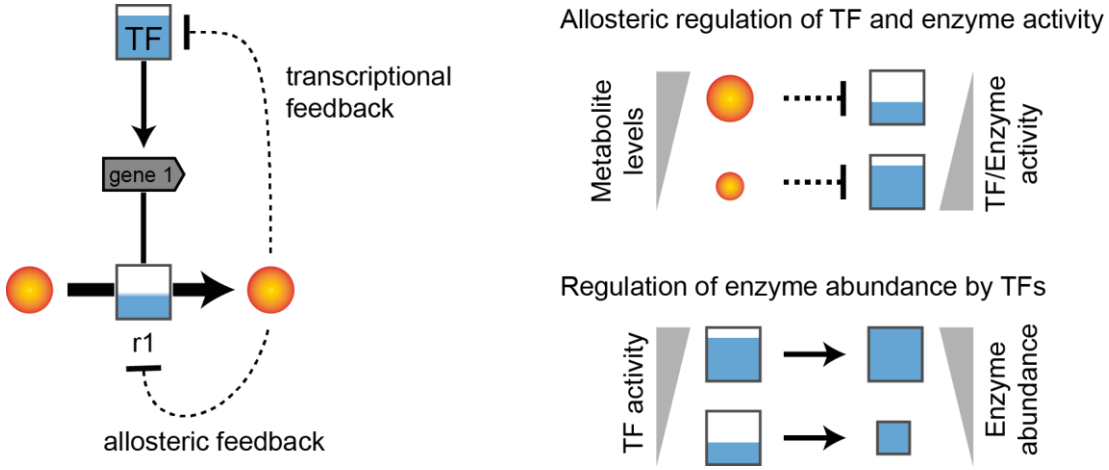


Figure 8. Allosteric regulation and transcriptional regulation*. Orange spheres are metabolites. The horizontal arrow is reaction 1 (r1) that is catalyzed by the enzyme E (blue square). The fill of E indicates the enzyme activity (high fill: high activity, low fill: low activity). The TF is a blue box and the fill of the box indicates TF activity (high fill: high activity, low fill: low activity). The dashed blunt arrows indicate negative allosteric and negative transcriptional feedback. The vertical pointed arrow indicates positive regulation of gene expression by the TF. Metabolite abundance affects enzyme and TF activity by allosteric and transcriptional feedback. The TF activity influences the enzyme abundance.*

To better understand crosstalk between metabolism and transcription, mechanistic mathematical models are required that integrate both layers. There are many mathematical models that either describe metabolism or the transcription regulation network. However, there are only few mathematical models that mechanistically link both layers[36,79,91].

Early models that combined metabolism and transcriptional regulation were developed based on the biochemical systems theory approach that uses the S-system formalism to approximate mechanistic rate laws[69]. For example, the autogenous regulation of a gene expression operon was compared to the inducible regulation approach regarding stability, robustness, and substrate overshoot. The model consisted of three layers and there was feedback between metabolism and transcription, and feedback between translation and transcription. The model predicted that the autogenous mechanism scores better when the system is controlled by a repressor. The opposite was predicted when the system is controlled by an activator[92].

A model that combined metabolism and transcription was used to study the temporal order of enzyme expression in unbranched pathways. Here, Michaelis-Menten kinetics[74] were used to model metabolic reactions and enzyme production rates. Enzyme production rates were balanced by dilution by growth. By optimizing enzyme production cost against the time to reach a target flux, the study concluded that sequential activation of transcription is optimal under these conditions in a linear pathway[91].

More recently, a mechanistic model of central carbon metabolism was built that combined metabolism and transcriptional regulation. Besides common metabolic reactions, it included known metabolite-TF interactions in central carbon metabolism. The transcription factors were assumed to directly modulate enzyme production rates. The model could reproduce the physiological behaviour of *E. coli*, including the metabolic adaptation from changing the carbon source from glucose to lactose. The study concluded that this was only possible due to metabolite-TF interactions that sense metabolic fluxes[79].

These results make clear that to understand metabolism, it is important to look at metabolism and enzyme-level regulation together. Modelling both layers may allow us to better understand properties and trade-offs that emerge from the interactions between metabolism and transcription. For example, it is still unclear how crosstalk between metabolism and transcription precisely determines enzyme abundances[28]. These models may help to identify unknown functional metabolite-TF interactions and elucidate their functional relevance at a genome scale, for example by using an ensemble modelling strategy[83].

1.4 Mass spectrometry-based metabolomics

1.4.1 Basics of mass spectrometry

The metabolome comprises all metabolites of *E. coli*. To understand metabolism, it is therefore necessary to measure all metabolites. Mass spectrometry-based metabolomics allows this by making snapshots of metabolite concentrations[93]. However, due to fast metabolite turnovers, low metabolite concentrations, and the condition dependent nature of metabolism, this is a challenging task[94,95]. Breakthroughs in mass spectrometry during the last decades have increased throughput and coverage of mass spectrometry-based metabolomics, allowing to rapidly probe large portions of the metabolome. Basically, a mass spectrometer is a balance that allows to determine the weights of ionized molecules. Conventional mass spectrometers consist of an ion source, a mass analyzer, and a detector[96].

Ion Source. In the ion source, sample molecules are electrically charged, allowing to measure their mass-to-charge ratio (*m/z*) in the mass spectrometer. One of the most frequently used ionization techniques is electrospray ionization[97] (ESI). ESI is a soft ionization technique that allows for the ionization of intact molecules. During ESI, the molecules are accumulating in small droplets on the tip of a heated capillary, and an electric field is applied. High temperatures cause droplets evaporation, which leads to the accumulation of charged molecules. When the Rayleigh limit is reached, the droplet becomes unstable, and the ions are dispersed into the vacuum of the mass analyzer[98]. Although ESI is considered a soft ionization technique, high temperatures and high voltages can cause undesired chemical reactions and fragmentations[99]. Moreover, ion suppression can be a problem in modern ESI sources. Ion suppression is a phenomenon in which compounds interfere with droplet formation in the ESI, affecting the number of charged molecules in the mass spectrometer[100].

Mass Analyzers. In the mass analyzer, the *m/z* is determined for each ion. The most frequently used mass analyzers are the Time-of-Flight[101] (ToF), the quadrupole[99], the orbitrap[102], and combinations thereof. In a ToF, ions are pulsed into the vacuum of a flight tube, in which they get accelerated by applying a voltage. The *m/z* of the ions can then be calculated from the time needed to reach the detector. Small molecules travel faster than large molecules. ToF mass analyzer offer a high mass resolution, and a mass

scale that is not limited by large masses[103]. ToF mass analyzer are often combined with quadrupole mass analyzers[104]. The quadrupole mass analyzer consists of four parallel metal rods. One pair of opposite rods uses radio frequency voltage, and the other pair of rods is applied with a direct current voltage. By modulating the ratio between radio frequency voltage and direct current voltage, only ions of a certain *m/z* are allowed to pass through the quadrupole. The other ions collide with the metal rods and are destroyed in the process. The quadrupole can act as an ion-guide, and as a mass filter, and often multiple quadrupoles are used[105].

Mass spectrometry-based metabolomics can be categorized into flow-injection mass spectrometry (FI-MS), in which the samples are directly injected into the mass spectrometer, and liquid chromatography-based tandem mass spectrometry (LC-MS/MS) that relies on an additional separation by liquid chromatography, and optional molecule fragmentation. Both types of MS analysis can be further categorized into targeted and untargeted analyses. In a targeted MS approach, only ion peaks corresponding to known metabolites are considered for peak annotation, whereas untargeted metabolomics considers all measured ion peaks[106].

1.4.2 Flow-injection mass spectrometry

In FI-MS, the samples are injected into the mobile phase that directly enters the mass spectrometer. Metabolites are then ionized in the ESI and distinguished solely by their *m/z* in the mass spectrum (**Figure 9a**). For peak annotation, *m/z* values of ion peaks are usually compared to reference databases like genome scale metabolic reconstructions[6,62]. FI-MS is faster than methods that use liquid chromatography, and it is therefore the method of choice to measure large amounts of samples in a short amount of time. FI-MS has been applied to measure thousands of strains or conditions in various organisms, ranging from *E. coli* to human cell lines[50,107–111]. While the speed of FI-MS is its biggest advantage, ion suppression and confidence in the annotated ion peaks can be problematic. Ion suppression is often caused by salt in the sample medium, and it is especially prominent in FI-MS, because the sample molecules enter the ion source at the same time[100,110].

Annotation confidence is limited by several effects. First, FI-MS does not allow us to distinguish between isomers and isobars in the MS1 spectrum. Therefore, single peaks are sometimes caused by multiple ions. Second, the number of ion peaks in the MS1

spectrum usually outweighs the number of known annotated metabolites. This means that there are either many unknown metabolites or that metabolites can produce multiple ion peaks. Thus, the origin of many ion peaks in the MS1 spectrum is still unclear, and it is also not clear, how these additional peaks affect peak annotation. Despite these shortcomings, FI-MS is a useful tool for screening large amounts of biological samples in a short amount of time.

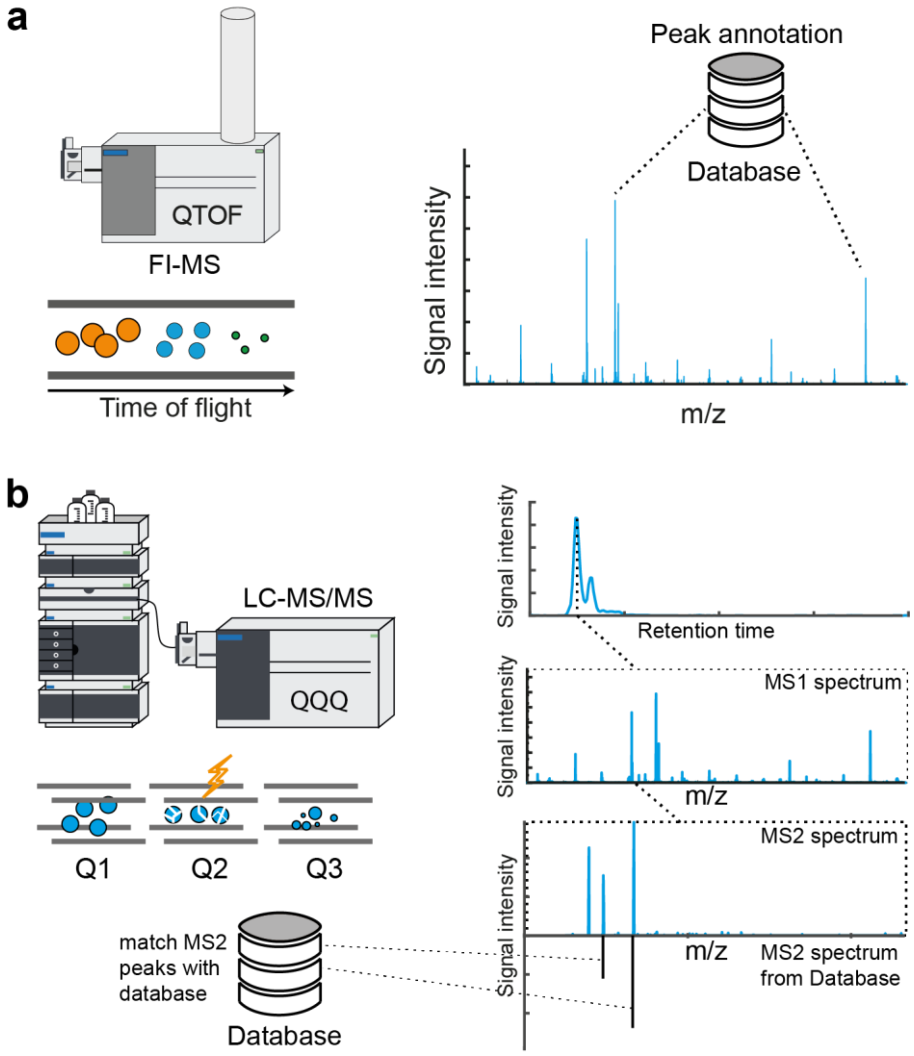


***Figure 9**. **Principles of flow-injection metabolomics and liquid chromatography-based tandem mass spectrometry. a) Flow-injection metabolomics (FI-MS)**. The samples enter the mass spectrometer directly with the mobile phase. Molecules of different masses have different sizes and colours. The molecules are separated in the Time-of-Flight tube based on their m/z. Light molecules are faster than heavy molecules. Peaks in the MS1 spectrum are then annotated by matching the m/z value to corresponding m/z values in a database. **b) Liquid chromatography-based tandem mass spectrometry (LC-MS/MS)**. The metabolites in the sample are injected into the LC, in which they are separated by physicochemical properties. The molecules (blue spheres) enter the triple-quadrupole mass spectrometer, consisting of three quadrupoles (Q1-Q3). Q1 is a mass filter, which filters out ions with specific masses (precursor mass). These ions enter Q2, in which they are fragmented by collision-induced dissociation (CID). The fragment masses of a chosen precursor mass are then filtered in Q3. For annotation, retention times from the LC, precursor masses, and fragment masses are usually matched with a spectral database containing such information (black spectrum). Credit to Martin Lempp for designing the mass spectrometers.*

1.4.3 Liquid chromatography-based tandem mass spectrometry

In liquid chromatography-based tandem mass spectrometry (LC-MS/MS), a liquid chromatography is used to first separate sample molecules according to their physicochemical properties. The molecules enter the ion source at different retention times. In the ion source, molecules are charged and dispersed into the vacuum of the mass analyzer. In targeted LC-MS/MS studies the mass analyzer often consists of three successive quadrupoles. In the first quadrupole (Q1), only ions with a defined *m/z* value (precursor mass) are retained. These ions enter the second quadrupole (Q2), in which they get fragmented by collision-induced dissociation (CID). During CID, inert gas (usually N2) is pulsed into the Q2 with a defined velocity (collision energy). Collision of the inert gas with sample ions causes fragmentation of ions[112]. The third quadrupole (Q3) can then filter for specific fragment masses. Sample metabolites can then be annotated based on their retention times, their precursor masses, and their fragment masses. Their identities are verified by matching the experimental spectra to compound specific information in databases (**Figure 9b**). For example, reference databases like the Metabolite and Chemical Entity Database (METLIN)[113], or the Human Metabolome Database[114] (HMDB) contain MS2 spectra for millions of compounds.

However, while these targeted approaches allow to annotate metabolites with high confidence, they depend on the completeness of the reference databases. For example, the METLIN database contains only 1 % of all the compounds listed in the PubChem database[113], and many other metabolites may be unknown, and not commercially available. Thus, increasing the metabolite coverage of these libraries will increase the scope of metabolites that can be confidently annotated. Moreover, in-source fragmentation is frequently observed in LC-MS/MS approaches, causing a much higher number of peaks than there are annotated metabolites[99,115,116]. Effective annotation of these peaks often requires a combination of chromatographic peak shape correlation analysis[93,117], MS2 information[93,117,118], and stable isotope labeled samples or standards[119,120]. In contrast to FI-MS, ion suppression is mitigated in LC-MS/MS approaches, since an LC separates many salt ions from the sample metabolites. At the same time, metabolites are also more likely to get lost during LC, making it more difficult to identify low abundant metabolites.

In conclusion, the metabolome can be measured by two complementary mass spectrometry-based methods, FI-MS and LC-MS/MS. FI-MS is fast and the method of choice for the screening of large numbers of samples. However, to obtain peak annotations with high confidence, an LC-MS/MS-based approach is more suitable. Both approaches suffer from high number of unknown peaks in the MS1 spectrum that can cause false-positive annotations. To obtain metabolite data of the necessary quality to answer biological questions, it is therefore crucial to understand the origins of these peaks.

1.5 Goal of this thesis

Metabolism of *E. coli* is complex, and to achieve a deeper understanding of metabolism, it is necessary to study systems behaviour and design principles of metabolism that go beyond single physical interactions. One of the most important design principles that governs molecular processes is robustness. Robustness allows cells to function despite perturbations, and one means to achieve robustness is feedback control executed by functional metabolite-protein interactions. These interactions allow the cell to modulate enzyme activity, and enzyme abundance. Since the complexity of these interactions prohibits an intuitive understanding of metabolism, high-quality multi-omics data need to be interpreted with the help of mechanistic mathematical models. However, most mechanistic mathematical models focus either on metabolism or transcription, and only few mechanistic models that combine both layers exist.

In the first part of this thesis (**Chapter 2** – **Chapter 5, Figure 10 yellow**), we developed mechanistic mathematical models that are regulated by either allosteric regulation, transcriptional regulation, or both mechanisms. Subsequently, we functionally characterize the impact of the regulations on the robustness of the systems, and potential trade-offs between different feedback mechanisms.

Besides a limited understanding of the crosstalk between metabolism and transcriptional regulation, recent studies suggest that there are likely many more unknown metabolite-TF interactions. Thus, in the second part of this thesis (**Chapter 6, Figure 10, purple**), we employ proteomics and metabolomics data from 283 genetic

knockdowns, corresponding to metabolic genes from a genome-scale model of *E. coli*, to map feedback regulation in *E. coli* metabolism.

Finally, a deep understanding of metabolism can only be achieved if the underlying metabolite data are reliable and comprehensive. Therefore, we analyse the flow-injection mass spectrometry approach in **Chapter 7 (Figure 10, blue)**, and we develop a method that is based on genetic knockdowns to measure low abundant metabolites in **Chapter 8**.

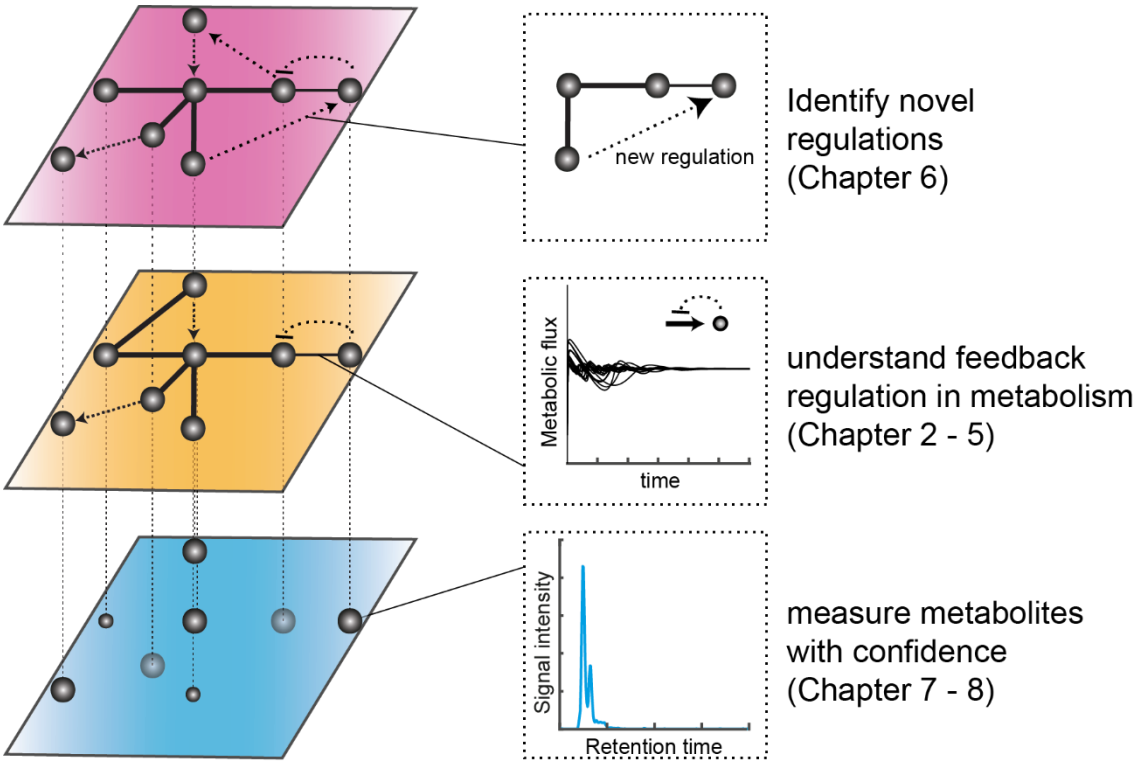


Figure 10. Thesis goals. *Spheres are metabolites. Solid lines indicate stoichiometric connections. Dashed arrows are regulatory interactions. (Blue) spheres are metabolites. Transparent spheres are metabolites whose identities are not certain. Small spheres are low abundant metabolites. The peak corresponds to a metabolite that can be measured. (Orange) Metabolites are connected by network stoichiometry and feedback regulation. Feedback regulation are dashed arrows. The dashed box shows a feedback regulated reaction that stabilizes into steady state. (Purple) New dashed arrows are new feedback regulations.*

1.6 References

(1) Ralser, M.; Varma, S. J.; Notebaart, R. A. The Evolution of the Metabolic Network over Long Timelines. *Current Opinion in Systems Biology* **2021**, *28*, 100402. https://doi.org/10.1016/j.coisb.2021.100402.

(2) Nielsen, J. Systems Biology of Metabolism. *Annu Rev Biochem* **2017**, *86*, 245–275. https://doi.org/10.1146/annurev-biochem-061516-044757.

(3) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563. https://doi.org/10.1038/227561a0.
(4) You, C.; Okano, H.; Hui, S.; Zhang, Z.; Kim, M.; Gunderson, C. W.; Wang, Y.-P.; Lenz, P.; Yan, D.; Hwa, T. Coordination of Bacterial Proteome with Metabolism by Cyclic AMP Signalling. *Nature* **2013**, *500* (7462), 301–306. https://doi.org/10.1038/nature12446.
(5) Kochanowski, K.; Okano, H.; Patsalo, V.; Williamson, J.; Sauer, U.; Hwa, T. Global Coordination of Metabolic Pathways in Escherichia Coli by Active and Passive Regulation. *Mol Syst Biol* **2021**, *17* (4), e10064. https://doi.org/10.15252/msb.202010064.
(6) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; Feist, A. M.; Palsson, B. O. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat Biotechnol* **2017**, *35* (10), 904–908. https://doi.org/10.1038/nbt.3956.
(7) Chubukov, V.; Gerosa, L.; Kochanowski, K.; Sauer, U. Coordination of Microbial Metabolism. *Nat Rev Microbiol* **2014**, *12* (5), 327–340. https://doi.org/10.1038/nrmicro3238.
(8) Bailey, J. E. Toward a Science of Metabolic Engineering. *Science* **1991**, *252* (5013), 1668–1675. https://doi.org/10.1126/science.2047876.
(9) Navid, A.; Ghim, C.-M.; Fenley, A. T.; Yoon, S.; Lee, S.; Almaas, E. Systems Biology of Microbial Communities. *Methods Mol Biol* **2009**, *500*, 469–494. https://doi.org/10.1007/978-1-59745-525-1_16.
(10) Gomes, A. P.; Ilter, D.; Low, V.; Drapela, S.; Schild, T.; Mullarky, E.; Han, J.; Elia, I.; Broekaert, D.; Rosenzweig, A.; Nagiec, M.; Nunes, J. B.; Schaffer, B. E.; Mutvei, A. P.; Asara, J. M.; Cantley, L. C.; Fendt, S.-M.; Blenis, J. Altered Propionate Metabolism Contributes to Tumour Progression and Aggressiveness. *Nat Metab* **2022**, *4* (4), 435–443. https://doi.org/10.1038/s42255-022-00553-5.
(11) Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F. J.; Doyle, J. Robustness of Cellular Functions. *Cell* **2004**, *118* (6), 675–685. https://doi.org/10.1016/j.cell.2004.09.008.
(12) Kitano, H. Biological Robustness. *Nat Rev Genet* **2004**, *5* (11), 826–837. https://doi.org/10.1038/nrg1471.
(13) Kitano, H. Towards a Theory of Biological Robustness. *Mol Syst Biol* **2007**, *3*, 137. https://doi.org/10.1038/msb4100179.
(14) Grimbs, S.; Selbig, J.; Bulik, S.; Holzhütter, H.-G.; Steuer, R. The Stability and Robustness of Metabolic States: Identifying Stabilizing Sites in Metabolic Networks. *Mol Syst Biol* **2007**, *3*, 146. https://doi.org/10.1038/msb4100186.
(15) Babul, J. Phosphofructokinases from Escherichia Coli. Purification and Characterization of the Nonallosteric Isozyme. *J Biol Chem* **1978**, *253* (12), 4350–4355.
(16) Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W. From Molecular to Modular Cell Biology. *Nature* **1999**, *402* (6761 Suppl), C47-52. https://doi.org/10.1038/35011540.
(17) Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. *Science* **2002**, *298* (5594), 824–827. https://doi.org/10.1126/science.298.5594.824.

(18) Kitano, H.; Oda, K. Robustness Trade-Offs and Host-Microbial Symbiosis in the Immune System. *Mol Syst Biol* **2006**, *2*, 2006.0022. https://doi.org/10.1038/msb4100039.
(19) Csete, M. E.; Doyle, J. C. Reverse Engineering of Biological Complexity. *Science* **2002**, *295* (5560), 1664–1669. https://doi.org/10.1126/science.1069981.
(20) Carlson, J. M.; Doyle, J. Highly Optimized Tolerance: Robustness and Design in Complex Systems. *Phys. Rev. Lett.* **2000**, *84* (11), 2529–2532. https://doi.org/10.1103/PhysRevLett.84.2529.
(21) Carlson, J. M.; Doyle, J. Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems. *Phys. Rev. E* **1999**, *60* (2), 1412–1427. https://doi.org/10.1103/PhysRevE.60.1412.
(22) Kitano, H. Cancer as a Robust System: Implications for Anticancer Therapy. *Nat Rev Cancer* **2004**, *4* (3), 227–235. https://doi.org/10.1038/nrc1300.
(23) Miele, L.; Evans, R. M. L.; Azaele, S. Redundancy-Selection Trade-off in Phenotype-Structured Populations. *Journal of Theoretical Biology* **2021**, *531*, 110884. https://doi.org/10.1016/j.jtbi.2021.110884.
(24) Monod, J.; Changeux, J. P.; Jacob, F. Allosteric Proteins and Cellular Control Systems. *J Mol Biol* **1963**, *6*, 306–329. https://doi.org/10.1016/s0022-2836(63)80091-1.
(25) Sauer, U. Metabolic Networks in Motion: 13C-Based Flux Analysis. *Mol Syst Biol* **2006**, *2*, 62. https://doi.org/10.1038/msb4100109.
(26) Fishov, I.; Zaritsky, A.; Grover, N. B. On Microbial States of Growth. *Mol Microbiol* **1995**, *15* (5), 789–794. https://doi.org/10.1111/j.1365-2958.1995.tb02349.x.
(27) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metabolic Engineering* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.
(28) Donati, S.; Sander, T.; Link, H. Crosstalk between Transcription and Metabolism: How Much Enzyme Is Enough for a Cell? *Wiley Interdiscip Rev Syst Biol Med* **2018**, *10* (1). https://doi.org/10.1002/wsbm.1396.
(29) Alam, M. T.; Olin-Sandoval, V.; Stincone, A.; Keller, M. A.; Zelezniak, A.; Luisi, B. F.; Ralser, M. The Self-Inhibitory Nature of Metabolic Networks and Its Alleviation through Compartmentalization. *Nat Commun* **2017**, *8*, 16018. https://doi.org/10.1038/ncomms16018.
(30) Macek, B.; Forchhammer, K.; Hardouin, J.; Weber-Ban, E.; Grangeasse, C.; Mijakovic, I. Protein Post-Translational Modifications in Bacteria. *Nat Rev Microbiol* **2019**, *17* (11), 651–664. https://doi.org/10.1038/s41579-019-0243-0.
(31) Shen-Orr, S. S.; Milo, R.; Mangan, S.; Alon, U. Network Motifs in the Transcriptional Regulation Network of Escherichia Coli. *Nat Genet* **2002**, *31* (1), 64–68. https://doi.org/10.1038/ng881.
(32) Vitreschak, A. G.; Lyubetskaya, E. V.; Shirshin, M. A.; Gelfand, M. S.; Lyubetsky, V. A. Attenuation Regulation of Amino Acid Biosynthetic Operons in Proteobacteria: Comparative Genomics Analysis. *FEMS Microbiol Lett* **2004**, *234* (2), 357–370. https://doi.org/10.1016/j.femsle.2004.04.005.
(33) Wang, T.; Simmel, F. C. Riboswitch-Inspired Toehold Riboregulators for Gene Regulation in Escherichia Coli. *Nucleic Acids Res* **2022**, gkac275. https://doi.org/10.1093/nar/gkac275.

(34) Scott, M.; Klumpp, S.; Mateescu, E. M.; Hwa, T. Emergence of Robust Growth Laws from Optimal Regulation of Ribosome Synthesis. *Mol Syst Biol* **2014**, *10*, 747. https://doi.org/10.15252/msb.20145379.

(35) Ma, H.-W.; Kumar, B.; Ditges, U.; Gunzer, F.; Buer, J.; Zeng, A.-P. An Extended Transcriptional Regulatory Network of Escherichia Coli and Analysis of Its Hierarchical Structure and Network Motifs. *Nucleic Acids Research* **2004**, *32* (22), 6643–6649. https://doi.org/10.1093/nar/gkh1009.

(36) Gerosa, L.; Kochanowski, K.; Heinemann, M.; Sauer, U. Dissecting Specific and Global Transcriptional Regulation of Bacterial Gene Expression. *Mol Syst Biol* **2013**, *9*, 658. https://doi.org/10.1038/msb.2013.14.

(37) Kochanowski, K.; Volkmer, B.; Gerosa, L.; Haverkorn van Rijsewijk, B. R.; Schmidt, A.; Heinemann, M. Functioning of a Metabolic Flux Sensor in Escherichia Coli. *Proc Natl Acad Sci U S A* **2013**, *110* (3), 1130–1135. https://doi.org/10.1073/pnas.1202582110.

(38) Zeng, J.; Hong, Y.; Zhao, N.; Liu, Q.; Zhu, W.; Xiao, L.; Wang, W.; Chen, M.; Hong, S.; Wu, L.; Xue, Y.; Wang, D.; Niu, J.; Drlica, K.; Zhao, X. A Broadly Applicable, Stress-Mediated Bacterial Death Pathway Regulated by the Phosphotransferase System (PTS) and the CAMP-Crp Cascade. *Proc Natl Acad Sci U S A* **2022**, *119* (23), e2118566119. https://doi.org/10.1073/pnas.2118566119.

(39) Kochanowski, K.; Sauer, U.; Chubukov, V. Somewhat in Control--the Role of Transcription in Regulating Microbial Metabolic Fluxes. *Curr Opin Biotechnol* **2013**, *24* (6), 987–993. https://doi.org/10.1016/j.copbio.2013.03.014.

(40) Diether, M.; Sauer, U. Towards Detecting Regulatory Protein–Metabolite Interactions. *Current Opinion in Microbiology* **2017**, *39*, 16–23. https://doi.org/10.1016/j.mib.2017.07.006.

(41) Umbarger, H. E. Evidence for a Negative-Feedback Mechanism in the Biosynthesis of Isoleucine. *Science* **1956**, *123* (3202), 848–848. https://doi.org/10.1126/science.123.3202.848.a.

(42) Dieckmann, R.; Pavela-Vrancic, M.; von Döhren, H.; Kleinkauf, H. Probing the Domain Structure and Ligand-Induced Conformational Changes by Limited Proteolysis of Tyrocidine Synthetase 1. *J Mol Biol* **1999**, *288* (1), 129–140. https://doi.org/10.1006/jmbi.1999.2671.

(43) Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; de Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P. Global Analysis of Protein Structural Changes in Complex Proteomes. *Nat Biotechnol* **2014**, *32* (10), 1036–1044. https://doi.org/10.1038/nbt.2999.

(44) Piazza, I.; Kochanowski, K.; Cappelletti, V.; Fuhrer, T.; Noor, E.; Sauer, U.; Picotti, P. A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **2018**, *172* (1–2), 358-372.e23. https://doi.org/10.1016/j.cell.2017.12.006.

(45) Reinhard, F. B. M.; Eberhard, D.; Werner, T.; Franken, H.; Childs, D.; Doce, C.; Savitski, M. F.; Huber, W.; Bantscheff, M.; Savitski, M. M.; Drewes, G. Thermal Proteome Profiling Monitors Ligand Interactions with Cellular Membrane Proteins. *Nat Methods* **2015**, *12* (12), 1129–1131. https://doi.org/10.1038/nmeth.3652.

(46) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Martinez Molina, D.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. Tracking Cancer Drugs in Living Cells by

Thermal Profiling of the Proteome. *Science* **2014**, *346* (6205), 1255784. https://doi.org/10.1126/science.1255784.

(47) Nikolaev, Y. V.; Kochanowski, K.; Link, H.; Sauer, U.; Allain, F. H.-T. Systematic Identification of Protein-Metabolite Interactions in Complex Metabolite Mixtures by Ligand-Detected Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2016**, *55* (18), 2590–2600. https://doi.org/10.1021/acs.biochem.5b01291.

(48) Hackett, S. R.; Zanotelli, V. R. T.; Xu, W.; Goya, J.; Park, J. O.; Perlman, D. H.; Gibney, P. A.; Botstein, D.; Storey, J. D.; Rabinowitz, J. D. Systems-Level Analysis of Mechanisms Regulating Yeast Metabolic Flux. *Science* **2016**, *354* (6311), aaf2786. https://doi.org/10.1126/science.aaf2786.

(49) Gerosa, L.; Haverkorn van Rijsewijk, B. R. B.; Christodoulou, D.; Kochanowski, K.; Schmidt, T. S. B.; Noor, E.; Sauer, U. Pseudo-Transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Syst* **2015**, *1* (4), 270–282. https://doi.org/10.1016/j.cels.2015.09.008.

(50) Fuhrer, T.; Zampieri, M.; Sévin, D. C.; Sauer, U.; Zamboni, N. Genomewide Landscape of Gene-Metabolome Associations in Escherichia Coli. *Mol Syst Biol* **2017**, *13* (1), 907. https://doi.org/10.15252/msb.20167150.

(51) Kochanowski, K.; Gerosa, L.; Brunner, S. F.; Christodoulou, D.; Nikolaev, Y. V.; Sauer, U. Few Regulatory Metabolites Coordinate Expression of Central Metabolic Genes in Escherichia Coli. *Mol Syst Biol* **2017**, *13* (1), 903. https://doi.org/10.15252/msb.20167402.

(52) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.

(53) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821. https://doi.org/10.1126/science.1225829.

(54) Telenti, A.; Pierce, L. C. T.; Biggs, W. H.; di Iulio, J.; Wong, E. H. M.; Fabani, M. M.; Kirkness, E. F.; Moustafa, A.; Shah, N.; Xie, C.; Brewerton, S. C.; Bulsara, N.; Garner, C.; Metzker, G.; Sandoval, E.; Perkins, B. A.; Och, F. J.; Turpaz, Y.; Venter, J. C. Deep Sequencing of 10,000 Human Genomes. *Proc Natl Acad Sci U S A* **2016**, *113* (42), 11901–11906. https://doi.org/10.1073/pnas.1613365113.

(55) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198–207. https://doi.org/10.1038/nature01511.

(56) Kitano, H. Systems Biology: A Brief Overview. *Science* **2002**, *295* (5560), 1662–1664. https://doi.org/10.1126/science.1069492.

(57) Wolkenhauer, O. Why Model? *Front Physiol* **2014**, *5*, 21. https://doi.org/10.3389/fphys.2014.00021.

(58) Bailey, J. E. Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnology Progress* **1998**, *14* (1), 8–20. https://doi.org/10.1021/bp9701269.

(59) Kremling, A.; Bettenbrock, K.; Gilles, E. D. A Feed-Forward Loop Guarantees Robust Behavior in Escherichia Coli Carbohydrate Uptake. *Bioinformatics* **2008**, *24* (5), 704–710. https://doi.org/10.1093/bioinformatics/btn010.

(60) Link, H.; Christodoulou, D.; Sauer, U. Advancing Metabolic Models with Kinetic Information. *Curr Opin Biotechnol* **2014**, *29*, 8–14. https://doi.org/10.1016/j.copbio.2014.01.015.
(61) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat Biotechnol* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.
(62) Orth, J. D.; Conrad, T. M.; Na, J.; Lerman, J. A.; Nam, H.; Feist, A. M.; Palsson, B. Ø. A Comprehensive Genome-Scale Reconstruction of Escherichia Coli Metabolism--2011. *Mol Syst Biol* **2011**, *7*, 535. https://doi.org/10.1038/msb.2011.65.
(63) Schuetz, R.; Kuepfer, L.; Sauer, U. Systematic Evaluation of Objective Functions for Predicting Intracellular Fluxes in Escherichia Coli. *Mol Syst Biol* **2007**, *3*, 119. https://doi.org/10.1038/msb4100162.
(64) Varma, A.; Palsson, B. O. Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type Escherichia Coli W3110. *Appl Environ Microbiol* **1994**, *60* (10), 3724–3731. https://doi.org/10.1128/aem.60.10.3724-3731.1994.
(65) Gatto, F.; Miess, H.; Schulze, A.; Nielsen, J. Flux Balance Analysis Predicts Essential Genes in Clear Cell Renal Cell Carcinoma Metabolism. *Sci Rep* **2015**, *5* (1), 10738. https://doi.org/10.1038/srep10738.
(66) Park, J. M.; Kim, T. Y.; Lee, S. Y. Prediction of Metabolic Fluxes by Incorporating Genomic Context and Flux-Converging Pattern Analyses. *Proceedings of the National Academy of Sciences* **2010**, *107* (33), 14931–14936. https://doi.org/10.1073/pnas.1003740107.
(67) Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **2010**, *330* (6007), 1099–1102. https://doi.org/10.1126/science.1192588.
(68) Scott, M.; Hwa, T. Bacterial Growth Laws and Their Applications. *Curr Opin Biotechnol* **2011**, *22* (4), 559–565. https://doi.org/10.1016/j.copbio.2011.04.014.
(69) Savageau, M. A. Introduction to S-Systems and the Underlying Power-Law Formalism. *Mathematical and Computer Modelling* **1988**, *11*, 546–551. https://doi.org/10.1016/0895-7177(88)90553-5.
(70) Liebermeister, W.; Klipp, E. Bringing Metabolic Networks to Life: Convenience Rate Law and Thermodynamic Constraints. *Theor Biol Med Model* **2006**, *3*, 41. https://doi.org/10.1186/1742-4682-3-41.
(71) Tran, L. M.; Rizk, M. L.; Liao, J. C. Ensemble Modeling of Metabolic Networks. *Biophys J* **2008**, *95* (12), 5606–5617. https://doi.org/10.1529/biophysj.108.135442.
(72) Voit, E. O.; Martens, H. A.; Omholt, S. W. 150 Years of the Mass Action Law. *PLoS Comput Biol* **2015**, *11* (1), e1004012. https://doi.org/10.1371/journal.pcbi.1004012.
(73) Michaelis, L.; Menten, M. L.; Johnson, K. A.; Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry* **2011**, *50* (39), 8264–8269. https://doi.org/10.1021/bi201284u.
(74) Briggs, G. E.; Haldane, J. B. A Note on the Kinetics of Enzyme Action. *Biochem J* **1925**, *19* (2), 338–339. https://doi.org/10.1042/bj0190338.
(75) Khodayari, A.; Maranas, C. D. A Genome-Scale Escherichia Coli Kinetic Metabolic Model k-Ecoli457 Satisfying Flux Data for Multiple Mutant Strains. *Nat Commun* **2016**, *7* (1), 13806. https://doi.org/10.1038/ncomms13806.

(76) García-Contreras, R.; Vos, P.; Westerhoff, H. V.; Boogerd, F. C. Why in Vivo May Not Equal in Vitro - New Effectors Revealed by Measurement of Enzymatic Activities under the Same in Vivo-like Assay Conditions. *FEBS J* **2012**, *279* (22), 4145–4159. https://doi.org/10.1111/febs.12007.

(77) Kaltenbach, H.-M.; Dimopoulos, S.; Stelling, J. Systems Analysis of Cellular Networks under Uncertainty. *FEBS Letters* **2009**, *583* (24), 3923–3930. https://doi.org/10.1016/j.febslet.2009.10.074.

(78) Gutenkunst, R. N.; Waterfall, J. J.; Casey, F. P.; Brown, K. S.; Myers, C. R.; Sethna, J. P. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Comput Biol* **2007**, *3* (10), 1871–1878. https://doi.org/10.1371/journal.pcbi.0030189.

(79) Kotte, O.; Zaugg, J. B.; Heinemann, M. Bacterial Adaptation through Distributed Sensing of Metabolic Fluxes. *Mol Syst Biol* **2010**, *6*, 355. https://doi.org/10.1038/msb.2010.10.

(80) Alves, R.; Savageau, M. A. Comparing Systemic Properties of Ensembles of Biological Networks by Graphical and Statistical Methods. *Bioinformatics* **2000**, *16* (6), 527–533. https://doi.org/10.1093/bioinformatics/16.6.527.

(81) Kuepfer, L.; Peter, M.; Sauer, U.; Stelling, J. Ensemble Modeling for Analysis of Cell Signaling Dynamics. *Nat Biotechnol* **2007**, *25* (9), 1001–1006. https://doi.org/10.1038/nbt1330.

(82) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metab Eng* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.

(83) Link, H.; Kochanowski, K.; Sauer, U. Systematic Identification of Allosteric Protein-Metabolite Interactions That Control Enzyme Activity in Vivo. *Nat Biotechnol* **2013**, *31* (4), 357–361. https://doi.org/10.1038/nbt.2489.

(84) Tsigkinopoulou, A.; Baker, S. M.; Breitling, R. Respectful Modeling: Addressing Uncertainty in Dynamic System Models for Molecular Biology. *Trends in Biotechnology* **2017**, *35* (6), 518–529. https://doi.org/10.1016/j.tibtech.2016.12.008.

(85) Fürtauer, L.; Nägele, T. Approximating the Stabilization of Cellular Metabolism by Compartmentalization. *Theory Biosci* **2016**, *135* (1–2), 73–87. https://doi.org/10.1007/s12064-016-0225-y.

(86) Kacser, H.; Burns, J. A. The Control of Flux. *Biochem Soc Trans* **1995**, *23* (2), 341–366. https://doi.org/10.1042/bst0230341.

(87) Chubukov, V.; Uhr, M.; Le Chat, L.; Kleijn, R. J.; Jules, M.; Link, H.; Aymerich, S.; Stelling, J.; Sauer, U. Transcriptional Regulation Is Insufficient to Explain Substrate-Induced Flux Changes in Bacillus Subtilis. *Mol Syst Biol* **2013**, *9*, 709. https://doi.org/10.1038/msb.2013.66.

(88) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli. *Nat Chem Biol* **2009**, *5* (8), 593–599. https://doi.org/10.1038/nchembio.186.

(89) Fendt, S.-M.; Buescher, J. M.; Rudroff, F.; Picotti, P.; Zamboni, N.; Sauer, U. Tradeoff between Enzyme and Metabolite Efficiency Maintains Metabolic Homeostasis upon Perturbations in Enzyme Capacity. *Mol Syst Biol* **2010**, *6*, 356. https://doi.org/10.1038/msb.2010.11.

(90) Christodoulou, D.; Link, H.; Fuhrer, T.; Kochanowski, K.; Gerosa, L.; Sauer, U. Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia Coli's Rapid Response to Oxidative Stress. *Cell Syst* **2018**, *6* (5), 569-578.e7. https://doi.org/10.1016/j.cels.2018.04.009.

(91) Zaslaver, A.; Mayo, A. E.; Rosenberg, R.; Bashkin, P.; Sberro, H.; Tsalyuk, M.; Surette, M. G.; Alon, U. Just-in-Time Transcription Program in Metabolic Pathways. *Nat Genet* **2004**, *36* (5), 486–491. https://doi.org/10.1038/ng1348.

(92) Savageau, M. A.; Voit, E. O. Recasting Nonlinear Differential Equations as S-Systems: A Canonical Nonlinear Form. *Mathematical Biosciences* **1987**, *87* (1), 83–115. https://doi.org/10.1016/0025-5564(87)90035-6.

(93) Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A. D.; Shen, Y.; Cowan, A.; McReynolds, M. R.; Kennedy, B. J.; Lato, A. M.; Campagna, S. R.; Singh, M.; Rabinowitz, J. D. Metabolite Discovery through Global Annotation of Untargeted Metabolomics Data. *Nat Methods* **2021**, *18* (11), 1377–1385. https://doi.org/10.1038/s41592-021-01303-3.

(94) Radoš, D.; Donati, S.; Lempp, M.; Rapp, J.; Link, H. Homeostasis of the Biosynthetic E. Coli Metabolome. *iScience* **2022**, *25* (7), 104503. https://doi.org/10.1016/j.isci.2022.104503.

(95) Riekeberg, E.; Powers, R. New Frontiers in Metabolomics: From Measurement to Insight. *F1000Res* **2017**, *6*, 1148. https://doi.org/10.12688/f1000research.11495.1.

(96) Wait, R. Introduction to Mass Spectrometry. *Methods Mol Biol* **1993**, *17*, 191–213. https://doi.org/10.1385/0-89603-215-9:191.

(97) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459. https://doi.org/10.1021/j150664a002.

(98) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9. https://doi.org/10.1021/ac302789c.

(99) Xue, J.; Domingo-Almenara, X.; Guijas, C.; Palermo, A.; Rinschen, M. M.; Isbell, J.; Benton, H. P.; Siuzdak, G. Enhanced In-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification. *Anal Chem* **2020**, *92* (8), 6051–6059. https://doi.org/10.1021/acs.analchem.0c00409.

(100) Müller, C.; Schäfer, P.; Störtzel, M.; Vogt, S.; Weinmann, W. Ion Suppression Effects in Liquid Chromatography-Electrospray-Ionisation Transport-Region Collision Induced Dissociation Mass Spectrometry with Different Serum Extraction Methods for Systematic Toxicological Analysis with Mass Spectra Libraries. *J Chromatogr B Analyt Technol Biomed Life Sci* **2002**, *773* (1), 47–52. https://doi.org/10.1016/s1570-0232(02)00142-3.

(101) Wolff, M. M.; Stephens, W. E. A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments* **1953**, *24* (8), 616–617. https://doi.org/10.1063/1.1770801.

(102) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: A New Mass Spectrometer. *J Mass Spectrom* **2005**, *40* (4), 430–443. https://doi.org/10.1002/jms.856.

(103) Boesl, U. Time-of-Flight Mass Spectrometry: Introduction to the Basics. *Mass Spectrom Rev* **2017**, *36* (1), 86–109. https://doi.org/10.1002/mas.21520.

(104) Miller, P. E.; Denton, M. B. The Quadrupole Mass Filter: Basic Operating Concepts. *J. Chem. Educ.* **1986**, *63* (7), 617. https://doi.org/10.1021/ed063p617.

(105) Li, C.; Chu, S.; Tan, S.; Yin, X.; Jiang, Y.; Dai, X.; Gong, X.; Fang, X.; Tian, D. Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Frontiers in Chemistry* **2021**, *9*.

(106) Zhang, X.; Zhu, X.; Wang, C.; Zhang, H.; Cai, Z. Non-Targeted and Targeted Metabolomics Approaches to Diagnosing Lung Cancer and Predicting Patient Prognosis. *Oncotarget* **2016**, *7* (39), 63437–63448. https://doi.org/10.18632/oncotarget.11521.

(107) Fuhrer, T.; Heer, D.; Begemann, B.; Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection-Time-of-Flight Mass Spectrometry. *Anal Chem* **2011**, *83* (18), 7074–7080. https://doi.org/10.1021/ac201267k.

(108) Beckmann, M.; Parker, D.; Enot, D. P.; Duval, E.; Draper, J. High-Throughput, Nontargeted Metabolite Fingerprinting Using Nominal Mass Flow Injection Electrospray Mass Spectrometry. *Nat Protoc* **2008**, *3* (3), 486–504. https://doi.org/10.1038/nprot.2007.500.

(109) Reiter, A.; Herbst, L.; Wiechert, W.; Oldiges, M. Need for Speed: Evaluation of Dilute and Shoot-Mass Spectrometry for Accelerated Metabolic Phenotyping in Bioprocess Development. *Anal Bioanal Chem* **2021**, *413* (12), 3253–3268. https://doi.org/10.1007/s00216-021-03261-3.

(110) Sarvin, B.; Lagziel, S.; Sarvin, N.; Mukha, D.; Kumar, P.; Aizenshtein, E.; Shlomi, T. Fast and Sensitive Flow-Injection Mass Spectrometry Metabolomics by Analyzing Sample-Specific Ion Distributions. *Nat Commun* **2020**, *11* (1), 3186. https://doi.org/10.1038/s41467-020-17026-6.

(111) Link, H.; Fuhrer, T.; Gerosa, L.; Zamboni, N.; Sauer, U. Real-Time Metabolome Profiling of the Metabolic Switch between Starvation and Growth. *Nat Methods* **2015**, *12* (11), 1091–1097. https://doi.org/10.1038/nmeth.3584.

(112) Wells, J. M.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol* **2005**, *402*, 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.

(113) Xue, J.; Guijas, C.; Benton, H. P.; Warth, B.; Siuzdak, G. METLIN MS2 Molecular Standards Database: A Broad Chemical and Biological Resource. *Nat Methods* **2020**, *17* (10), 953–954. https://doi.org/10.1038/s41592-020-0942-5.

(114) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res* **2007**, *35* (Database issue), D521-526. https://doi.org/10.1093/nar/gkl923.

(115) Senan, O.; Aguilar-Mogas, A.; Navarro, M.; Capellades, J.; Noon, L.; Burks, D.; Yanes, O.; Guimerà, R.; Sales-Pardo, M. CliqueMS: A Computational Tool for Annotating in-Source Metabolite Ions from LC-MS Untargeted Metabolomics Data Based on a Coelution Similarity Network. *Bioinformatics* **2019**, *35* (20), 4089–4097. https://doi.org/10.1093/bioinformatics/btz207.

(116) Guo, J.; Shen, S.; Xing, S.; Yu, H.; Huan, T. ISFrag: De Novo Recognition of In-Source Fragments for Liquid Chromatography-Mass Spectrometry Data. *Anal Chem* **2021**, *93* (29), 10243–10250. https://doi.org/10.1021/acs.analchem.1c01644.

(117) Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; Korf, A.; Pluskal, T.; Kameník, Z.; Jarmusch, A. K.; Caraballo-Rodríguez, A. M.; Weldon, K. C.; Nothias-Esposito, M.; Aksenov, A. A.; Bauermeister, A.; Albarracin Orio, A.; Grundmann, C. O.; Vargas, F.; Koester, I.; Gauglitz, J. M.; Gentry, E. C.; Hövelmann, Y.; Kalinina, S. A.; Pendergraft, M. A.; Panitchpakdi, M.; Tehan, R.; Le Gouellec, A.; Aleti, G.; Mannochio Russo, H.; Arndt, B.; Hübner, F.; Hayen, H.; Zhi, H.; Raffatellu, M.; Prather, K. A.; Aluwihare, L. I.; Böcker, S.; McPhail, K. L.; Humpf, H.-U.; Karst, U.; Dorrestein, P. C. Ion Identity Molecular Networking for Mass Spectrometry-Based Metabolomics in the GNPS Environment. *Nat Commun* **2021**, *12* (1), 3832. https://doi.org/10.1038/s41467-021-23953-9.

(118) Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M. A.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; Böcker, S. Systematic Classification of Unknown Metabolites Using High-Resolution Fragmentation Mass Spectra. *Nat Biotechnol* **2021**, *39* (4), 462–471. https://doi.org/10.1038/s41587-020-0740-8.

(119) Guder, J. C.; Schramm, T.; Sander, T.; Link, H. Time-Optimized Isotope Ratio LC-MS/MS for High-Throughput Quantification of Primary Metabolites. *Anal Chem* **2017**, *89* (3), 1624–1631. https://doi.org/10.1021/acs.analchem.6b03731.

(120) Hartl, J.; Kiefer, P.; Kaczmarczyk, A.; Mittelviefhaus, M.; Meyer, F.; Vonderach, T.; Hattendorf, B.; Jenal, U.; Vorholt, J. A. Untargeted Metabolomics Links Glutathione to Bacterial Cell Cycle Progression. *Nat Metab* **2020**, *2* (2), 153–166. https://doi.org/10.1038/s42255-019-0166-0.

Chapter 2 A mathematical model predicts a trade-off between robustness and efficiency in amino acid biosynthesis

Sander, T.; Farke, N.; Diehl, C.; Kuntz, M.; Glatter, T.; Link, H. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. Coli by Enforcing Enzyme Overabundance. Cell Syst 2019, 8 (1), 66-75. e8.
https://doi.org/10.1016/j.cels.2018.12.005

Parts of this chapter are published in Cell Systems 2019, 8 (1), 66-75. e8. My contribution to this work included the design of the mathematical models, and the analysis of the mathematical models. I performed literature research of kinetic parameters of allosteric and transcriptional feedback regulation. I co-wrote the manuscript.

2.1 Abstract

Microbes regularly face fluctuating environments and to ensure stable growth, they require a constant supply of amino acids. One means to achieve robustness is allosteric feedback control. In amino acid biosynthesis, the amino acid usually feedback inhibits the first committed step of its own biosynthesis. This way, the amino acid supply can be adjusted to the requirements of the cell. At the same time, amino acids can interact with transcription factors to control the abundance of pathway enzymes. This mechanism is crucial to ensure efficient use of resources. Although allosteric feedback and transcriptional feedback have been known for a long time, it is still unclear how they act in concert. In this chapter, we developed a mathematical model of amino acid biosynthesis that combines allosteric feedback inhibition and transcriptional feedback inhibition. The model showed that allosteric feedback increases robustness by increasing enzyme levels, and that transcriptional feedback increases efficiency by reducing enzyme levels, and that both mechanisms work together to solve a robustness-efficiency trade-off in amino acid biosynthesis. Moreover, model, and experimental data showed that wild type *E. coli* occupies the middle of a trade-off where both objectives are balanced. Thus, the model showed that both feedback mechanisms work together to finely tune enzyme levels to ensure robust, yet efficient amino acid biosynthesis in *E. coli*.

2.2 Introduction

Metabolism is regulated by many different mechanisms, and on several layers[1]. While allosteric feedback regulation controls enzyme activity, transcriptional regulation controls enzyme abundance. However, because the metabolic flux of reaction depends on enzyme activity and enzyme abundance, it becomes difficult to disentangle the contributions of both regulation mechanisms on individual reactions[2,3]. To study the interplay between allosteric feedback and transcriptional feedback, amino acid biosynthesis pathways are suitable model pathways since they are regulated by both mechanisms (**Figure 11**).

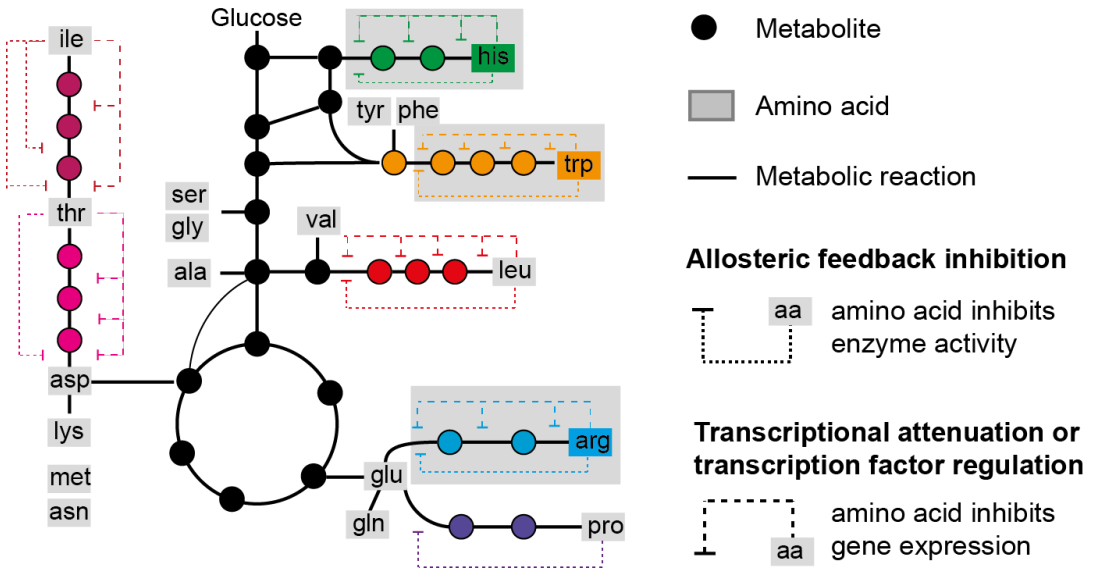


Figure 11. Metabolic network and architecture of amino acid biosynthesis in E. coli. *Simplified metabolic map of central carbon metabolism. Metabolites are spheres (black in catabolic pathways and coloured in amino acid biosynthesis pathways). Amino acids are squares. Metabolites and amino acids are connected by metabolic reactions (straight lines). The principle of feedback regulation (inhibition) is shown by dashed, blunt arrows. Tryptophan, arginine and histidine pathways are highlighted by grey boxes.*

16/20 amino acids allosterically inhibit the committed step of their own biosynthesis. This feedback inhibition ensures homeostasis of end products[4]. Removing allosteric feedback regulation was mainly studied *in-vitro*[5] and for the purpose of improving industrial production strains[6]. In nucleotide metabolism of *E. coli*, removing the allosteric feedback had no apparent effect on nucleotide levels[7]. In this case, other mechanisms ensured channelling excess of nucleotides into degradation pathways, showing how redundancy enables robust metabolism. Conversely, theoretical studies have shown that allosteric feedback inhibition is crucial to ensure end-product homeostasis[8], metabolic robustness[9], control of metabolic flux[10] and optimal growth[11,12].

Enzyme levels of amino acid biosynthesis are controlled via transcription factors (TFs) and transcriptional attenuation. Overall, four TFs (ArgR, TrpR, TyrR, Lrp) control enzyme levels in 19/20 amino acid pathways, ensuring that enzymes are only produced when they are needed[13,14]. However, recent data suggests that enzymes are often more abundant than necessary, and that such enzyme overabundance conveys robustness against internal and external perturbations[15,16].

Mathematical models have been used to study metabolism or transcriptional regulation. However, there are only few cases where mathematical models have been developed that employ both layers[14,17,18].

To study the interplay between allosteric feedback of enzyme activity and transcriptional feedback of enzyme abundance, we developed a mechanistic mathematical model of amino acid biosynthesis. Dysregulation of the model showed that allosteric feedback increases enzyme levels, and that transcriptional feedback decreases enzyme levels. To validate the model, we measured protein, metabolite and flux changes in mutants of the arginine, histidine, and tryptophan pathways lacking allosteric feedback and transcriptional feedback, respectively. The model predicted that allosteric feedback and transcriptional feedback solve a trade-off between robustness and efficiency, and that wild type *E. coli* occupies the middle of the trade-off, where both objectives are balanced. Thus, the model showed that both feedback mechanisms work together to finely tune enzyme levels to ensure robust, yet efficient amino acid biosynthesis in *E. coli*[12].

2.3 Allosteric feedback inhibition causes enzyme overabundance in amino acid biosynthesis

To understand the interplay between allosteric regulation of enzyme activity and transcriptional regulation of enzyme abundance, we created *E. coli* mutants lacking allosteric feedback (point mutation in the allosteric site of the enzyme) or transcriptional feedback (deletion of the respective transcription factor/attenuation leader peptide) in histidine, tryptophan and arginine biosynthesis pathways. This resulted in three *E. coli* mutants without allosteric feedback regulation (argA*, trpE*, hisG*) and three *E. coli* mutants without transcriptional feedback regulation (ΔargR, ΔtrpR, ΔhisL).

First, we measured metabolites, proteins and metabolic fluxes of the allosterically dysregulated strains (argA*, trpE*, hisG*), and compared them to the corresponding wild-type strain. The amino acids of the respective dysregulated strains showed the strongest change in all three strains. For example, arginine accumulated in the argA* strain, while the concentrations of the other amino acids remained stable (**Figure 12a**).

Thus, removing the allosteric feedback resulted in a local upregulation of the respective amino acid, indicating that allosteric feedback inhibition maintains low amino acid concentrations in amino acid biosynthesis.

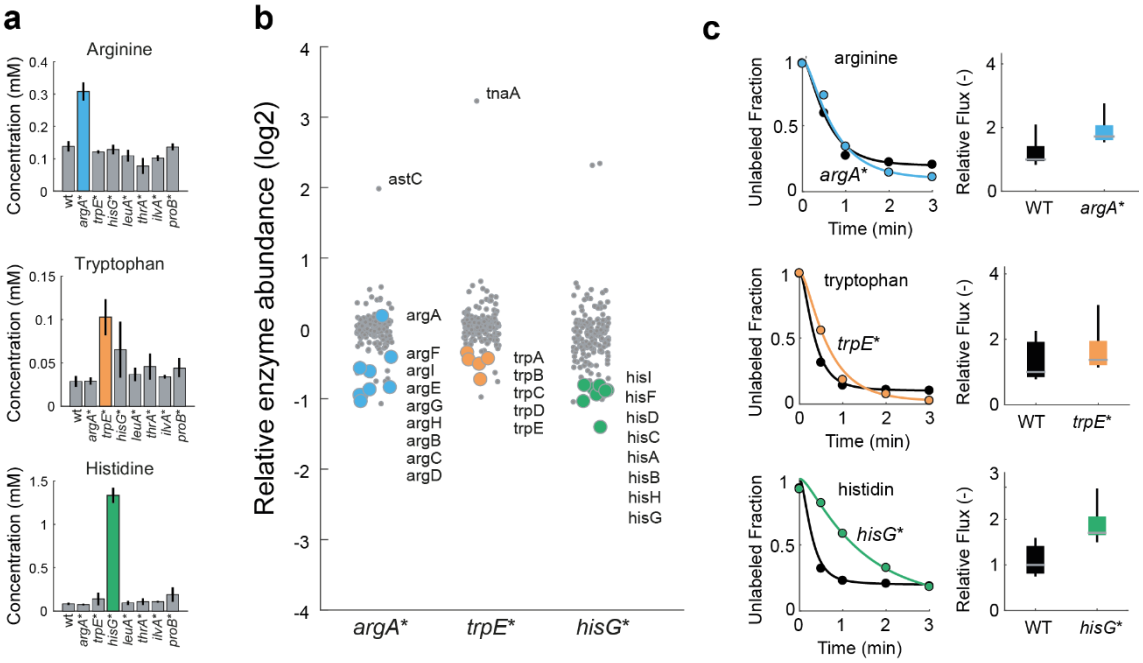


Figure 12. Metabolome, Proteome and fluxes of allosterically dysregulated arginine, tryptophan and histidine strains. a) Relative concentrations of intracellular amino acids in wild type *E. coli* and the seven dysregulated mutants, including argA*, trpE*, hisG*. Bar plots show absolute concentrations of the amino acid in the dysregulated pathways. Data are represented as mean, and error bars are ± SD (n = 3). **b)** Abundance of 173 enzymes in amino acid metabolism (out of 204 enzymes in total), relative to the level in the wild type. Data are represented as mean (n = 3). For each strain the enzymes in the dysregulated pathway are shown as colored dots. Enzymes in degradation pathways of arginine and tryptophan are indicated by their names. **c)** Decay of unlabeled amino acids in the wild-type *E. coli* (black) and the three dysregulated mutants. The measured amino acid is indicated above each graph. Cells were loaded from shake flasks onto filters and perfused with 15N-medium for different lengths of time (0, 30, 60, 120, and 180 seconds). Dots are means of n = 2 samples for each time point. Lines are means of 1,000 fits of decay rates based on equations for kinetic flux profiling. Box plots show fluxes based on the 1,000 fits, relative to the median flux estimate in the wild-type. Boxes contain 50% and whiskers 99% of the flux estimates. Parts of this figure are published in Cell Systems 2019, 8 (1), 66-75. e8.

We next inspected the proteome changes in the amino acid pathways and observed reduced enzyme levels in all three allosteric mutants compared to the wild-type strain (**Figure 12b**). This is likely because elevated amino acid levels activate the corresponding transcription factors, which downregulate the pathway enzymes. Thus, if the amino acid concentration is higher than needed, the cell reduces enzyme levels in the respective pathways. However, the reduction of enzyme levels did not cause a limitation in amino

acid flux (**Figure 12c**), suggesting that enzymes are not operating at their full capacity in wild type *E. coli* during growth on glucose, as indicated by previous studies[16,19,20].

We then hypothesized that allosteric feedback inhibition enables enzyme overabundance in amino acid biosynthesis by maintaining low levels of amino acids that cause a de-repression of pathway enzymes. To disentangle the mechanisms that cause enzyme overabundance, we next developed a mathematical model to study the interplay between allosteric feedback and transcriptional feedback (enzyme-level regulation).

2.4 Interplay of allosteric feedback inhibition and enzyme-level regulation

To better understand the interplay between allosteric feedback inhibition and transcriptional feedback inhibition, we developed a mechanistic mathematical model of amino acid metabolism that combines metabolism and enzyme-level regulation. The minimal model describes a two-step pathway that includes two enzymes (e1 and e2), and two metabolites (m1 and m2). The end-product m2 is an amino acid that is consumed in a third metabolic reaction for growth. Each enzyme (e1 and 2) is produced by a constant production rate and is consumed via dilution by growth.

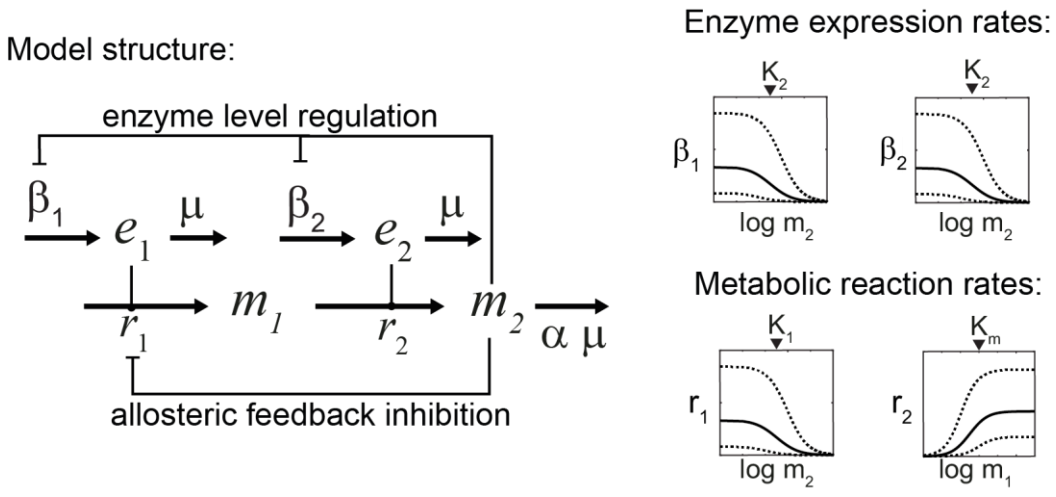


Figure 13. Stoichiometry and structure of the kinetic model. *m_1 and m_2 are metabolites, e_1 and e_2 are enzymes. Kinetics of the enzyme catalysed reactions r_1 and r_2, as well as kinetics of enzyme expression rates β_1 and β_2, are sampled in the indicated intervals. Blunt arrows are allosteric feedback inhibition or enzyme level regulation (transcriptional regulation). This figure is published in Cell Systems 2019, 8 (1), 66-75. e8.*

The amino acid m2 feedback inhibits the activity of the first enzyme in the pathway, as well as the production rates of e1 and e2. The enzyme production rates, and the first reaction follow simple inhibition kinetics. The second reaction follows Michaelis-Menten kinetics (**Figure 13**). Thus, this model is a simplified representation of amino acid metabolism that is controlled by allosteric feedback inhibition and by transcriptional feedback inhibition (enzyme-level regulation).

To analyse the model, we used an ensemble modelling approach. Therefore, we fixed the flux in the pathway to the amino acid requirement that corresponds to the growth rate of *E. coli* on glucose. We then created an ensemble of models by sampling seven kinetic parameters 5,000 times from physiologically relevant intervals based on literature values. For each of the 5,000 parameter sets, we calculated the steady state enzyme and metabolite concentrations for a complete model with allosteric feedback on enzyme activity and transcriptional feedback on enzyme abundance (complete model, gray in **Figure 14a**), for a model including only feedback on enzyme abundance (only transcriptional feedback model, blue in **Figure 14a**), and for a model including only feedback on enzyme activity (only allosteric feedback model, orange in **Figure 14a**).

The simulated concentrations of e1, e2, m1, and m2 qualitatively matched the measured protein and metabolite data (shown for the ΔargR, and argA* strain in **Figure 14b**): the two enzymes decreased in the only-transcriptional feedback model, whereas the end-product m2 increased. Moreover, the simulated concentration of the intermediate m1 matched the measured increase of intermediates in amino acid pathways. In the only-allosteric feedback model, the enzyme levels and the end-product m2 increased while the intermediate metabolite m1 decreased, matching the data obtained for the ΔargR strain (**Figure 14b**).

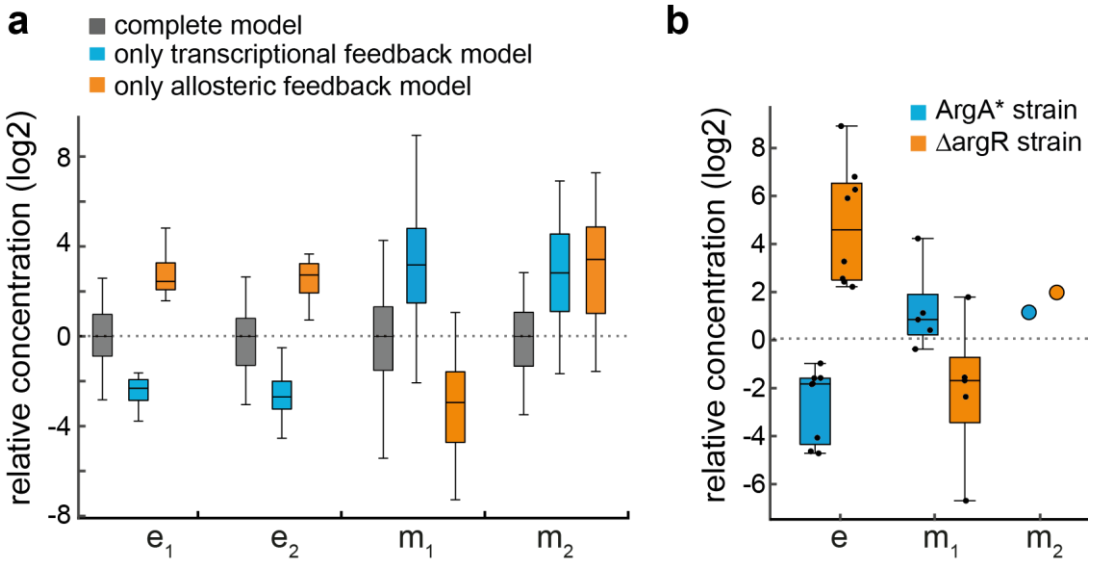


Figure 14. Simulated metabolite and enzyme changes match experimental changes. a) *Steady state concentrations of e_1, e_2, m_1 and m_2 calculated with 5000 random parameter sets for the complete model (grey), the model with only transcriptional regulation (blue) and the model with only allosteric regulation (orange). Boxes contain 50% and whiskers 99% of the simulated concentrations. All concentrations are normalized to the median concentrations of the complete model.* ***b)*** *Enzyme and metabolite changes for the argA* strain and the ⊿argR strain. Pathway enzymes are e, pathway metabolites are m1 and arginine is m2. The argA* strain is blue, the ⊿argR strain is orange. Boxes contain 50% and whiskers 99% of the data points. Black dots are experimental values of the pathway metabolites (mean, n=3). Arginine levels are shown as sphere (mean, n = 3). Parts of this figure are published in Cell Systems 2019, 8 (1), 66-75. e8*

Thus, a simple mechanistic model of amino acid biosynthesis confirmed our hypothesis that allosteric feedback inhibition causes enzyme overabundance by increasing enzyme levels, and that transcriptional regulation reduces enzyme levels. We then explored if other types of enzyme inhibition could cause a similar increase in enzyme expression by replacing the allosteric feedback in the model with competitive product inhibition of the second reaction (**Figure 15a**). However, removing competitive product inhibition was compensated by lower substrate concentrations (m1) and not by lower enzyme levels (**Figure 15b**). This indicated that enzyme overabundance does not emerge from all types of enzyme inhibition[12].

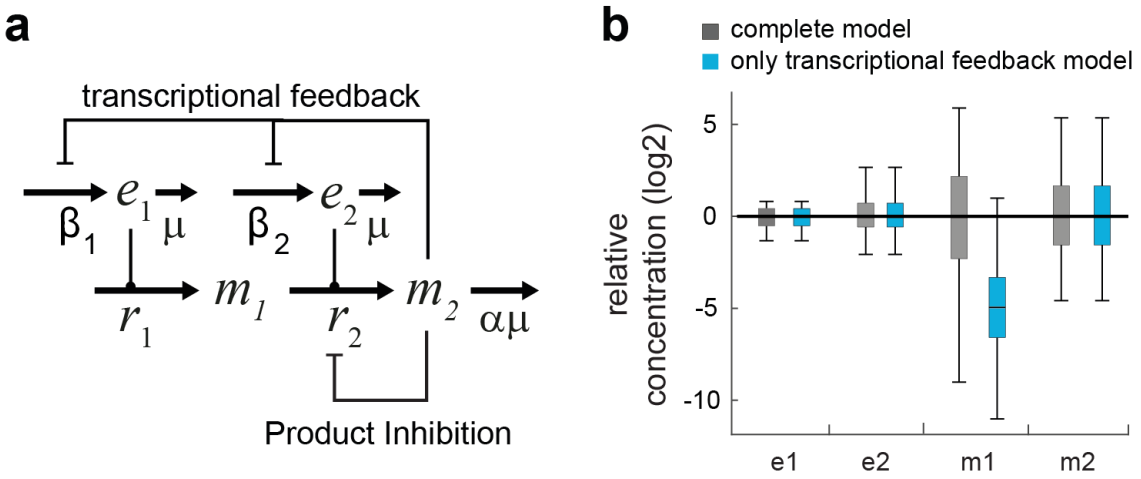


Figure 15. Mechanistic model with product inhibition (a) *Model Structure. Metabolite 2 inhibits reaction 2 and competes with Metabolite 1 for the active site of enzyme 2. Product inhibition was modelled using the following equation:* $r_2 = \beta_{12} \cdot e_2 \cdot \frac{m1}{m1+Km\cdot\left(1+\frac{m2}{K1}\right)}$ *.* ***b)*** *Steady state concentrations of* e_1, e_2 m_1 *and* m_2 *calculated with 5000 simulations for the complete model (grey), and the single feedback model (blue). Boxes contain 50% and whiskers 99% of the simulated concentrations. All concentrations are normalized to the median concentrations of the complete model. This figure is published in Cell Systems 2019, 8 (1), 66-75. e8.*

2.5 Allosteric feedback and transcriptional feedback balance robustness and efficiency in amino acid biosynthesis

Next, we investigated the function that emerges from the interplay between allosteric feedback inhibition of enzyme activity and transcriptional feedback inhibition of enzyme production rates. While keeping low enzyme levels is more efficient due to high enzyme cost, high enzyme levels could provide a cellular benefit by providing robustness against perturbations in enzyme expression. To test this with the mechanistic model, we employed a numerical parameter continuation method to quantify robustness[21]. This method iteratively varies a model parameter until a bifurcation point is reached (the model becomes unstable). Robustness can then be defined as the percentage change of this parameter that is tolerated. Using this method, we calculated robustness against perturbations of the maximal expression rate of the second enzyme ($\beta_{2,max}$) in the complete model with 5,000 randomly sampled parameter sets (**Figure 16**). Changing $\beta_{2,max}$ reflects genetic or environmental perturbations of gene expression that can cause a bottleneck in the pathway. In agreement with our expectations, models with high enzyme

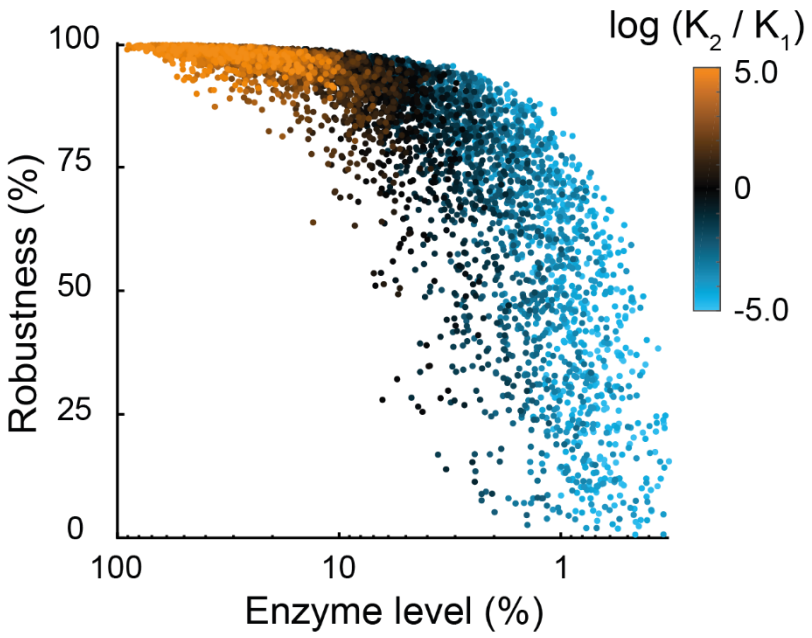


Figure 16. Robustness-Efficiency trade-off*. Enzyme levels (sum of e_1 and e_2) and robustness against perturbations of $\beta_{2,max}$, for 5000 simulations of the complete model (dots). The color of each dot shows the ratio of allosteric inhibition constant (K_1) and transcriptional inhibition constant (K_2) in the respective model. 100% robustness corresponds to the maximal possible downregulation of $\beta_{2,max}$. 100% enzyme abundance corresponds to the maximum theoretical enzyme concentration in the model. This figure is published in Cell Systems 2019, 8 (1), 66-75. e8.*

levels showed increased robustness, while models with lower enzyme levels were more sensitive to perturbations of enzyme expression (**Figure 16**). However, robustness was not proportional to the enzyme level: a relatively small increase of enzyme levels already conferred a large robustness benefit. Very high enzyme levels, in comparison, did not increase robustness substantially over more subtle changes in enzyme abundance. Our model thus reveals a trade-off between protein costs (efficiency) and robustness, which can be solved by sensitively balancing enzyme levels. The optimal balance of enzyme levels occurs in models occupying the middle of the trade-off. Here, the models have similarly strong feedback on enzyme activity and on enzyme abundance (indicated by similar inhibition constants Ki, black dots in **Figure 16**). This model result is supported by our experimental data of the argA*, trpE* and hisG* mutants (**Figure 12**), which demonstrate that wild-type *E. coli* does not operate with minimal enzyme levels in these pathways (blue dots in **Figure 16**).

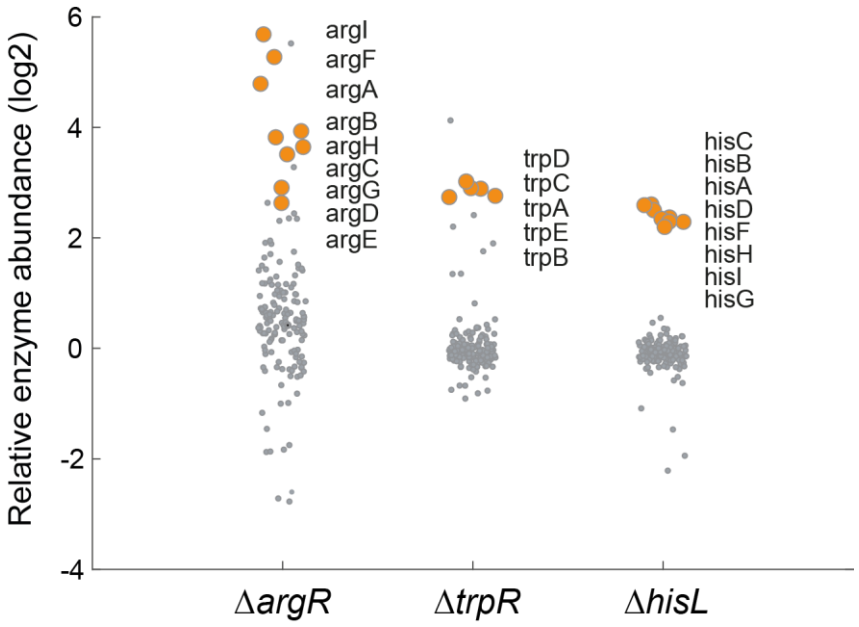


Figure 17. Enzyme levels of transcriptionally dysregulated strains*. Abundance of enzymes in amino acid metabolism of the ΔargR, ΔtrpR and ΔhisL mutants relative to the wild-type. Data are represented as mean (n = 3). For each strain the enzymes in the dysregulated pathway are shown as orange dots. This figure is published in Cell Systems 2019, 8 (1), 66-75. e8.*

In conclusion, analysis of the mathematical models and the dysregulated mutants allowed us to disentangle the regulatory interplay between regulation of enzyme activity (allosteric feedback) and regulation of enzyme abundance (transcriptional feedback) in the arginine, tryptophan, and histidine pathways. Removing transcriptional feedback regulation increased enzyme levels (**Figure 17**) and removing allosteric feedback regulation decreased enzyme levels (**Figure 12**). The model predicted that inhibition constants of the two feedback regulations must have similar magnitudes, if feedback on enzyme activity and enzyme abundance are simultaneously active (black dots in **Figure 16**). In the literature, binding constants for allosteric and transcriptional feedback have similar magnitudes, supporting the existence of a two-pronged regulation strategy. (**Appendix, Table 2).**

2.6 Oscillations in amino acid biosynthesis as by-product of robust efficiency

Thus, allosteric feedback and transcription feedback balance two conflicting objectives: robustness and efficiency. However, oscillations are often undesired by-products of the

hard trade-offs between robustness and efficiency[22]. To test if oscillations are also apparent in amino acid metabolism, we designed a smaller model with only one

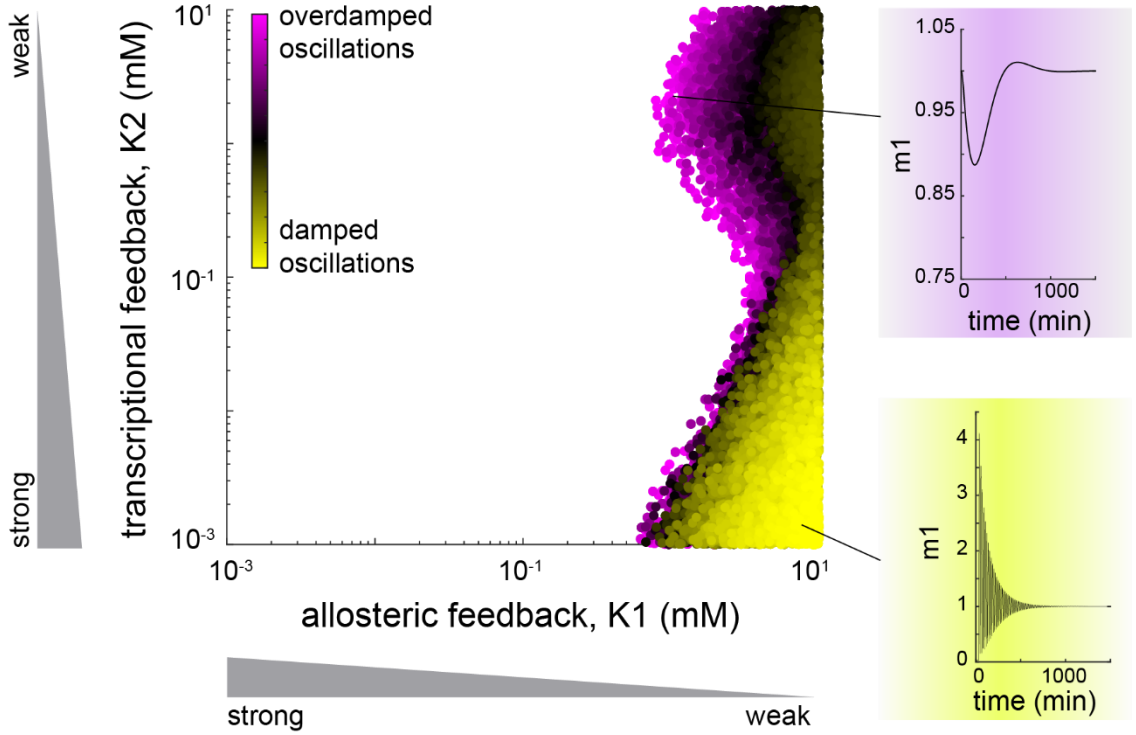


Figure 18. Oscillations in amino acid biosynthesis*. Scatterplot showing a trade-off between binding constants of allosteric feedback inhibition (K1) and transcriptional feedback inhibition (K2). A dot was made when the eigenvalues of the Jacobian matrix of the system in steady state had an imaginary part. Yellow dots are strong (damped) oscillations. Purple dots are overdamped oscillations. The time course simulation of m1 in the yellow box corresponds to a yellow dot at the indicated position. The time course simulation of m1 in the purple box corresponds to a purple dot at the indicated position. The enzyme expression rate β1 was reduced 20-fold at t = 30 min.*

metabolite (amino acid m1) and one enzyme. We simulated 5,000 parameter sets from physiologically relevant ranges and ensured that all models converged to stable steady states. We then wondered if the interplay between allosteric feedback and transcriptional feedback causes the model to produce oscillations of amino acids levels. To analyse if a system in steady state can produce oscillations upon perturbation, we inspected the eigenvalues of the Jacobian matrix. Eigenvalues consist of a real part and an imaginary part. A non-zero imaginary part indicates that the system can produce oscillations, while the magnitude of a negative real part indicates how strongly the system is damped. Thus, to identify oscillations, we looked at the ratio between the imaginary part and the real part of the eigenvalues of 5,000 models. For many combinations of allosteric feedback and transcriptional feedback, the model predicted

no oscillations (**Figure 18**, white space). However, when allosteric feedback was weak and transcriptional feedback was strong, the system showed damped oscillations (**Figure 18**, yellow dots, and yellow box with time course of the amino acid m1). This agrees with our observation that allosteric feedback increases robustness of the system. At weaker transcriptional feedback, the oscillations were overdamped (**Figure 18**, purple dots and purple box with time course of the amino acid m1). However, oscillations only occurred when both feedback regulations were equally weak (similar K-value). These results indicate that oscillations in amino acid biosynthesis are a by-product of robust efficiency, as was already shown for glycolysis[22].

2.7 Enzyme overabundance provides robustness against genetic perturbations

The model predicted that the allosterically dysregulated strains (ArgA*, hisG*, trpE*) are more susceptible against genetic perturbations than the wild-type strain. To test this experimentally, we used CRISPR interference (CRISPRi) to knockdown target genes in the respective amino acid biosynthesis pathways (*argE*, *hisB*, *trpA*). This experiment resembled the numerical reduction of the parameter $\beta_{2,max}$ in the model analysis. Perturbation of $\beta_{2,max}$ resulted in e2 being diluted by growth, causing a bottleneck in the pathway. Similarly, CRISPRi blocks the transcription of the target enzyme, leading to enzyme dilution by growth and consequently to a bottleneck in the pathway. We then perturbed all pairwise combinations of CRISPRi targets and dysregulated strains, and measured growth rates in all strains (**Figure 19**). We observed the largest growth defects (more than 50% reduction of the growth rate) when CRISPRi was used in the pathways of the corresponding dysregulated strains (i.e. argE in argA*, hisB in hisG* and trpA in trpE*). These data confirm that the dysregulated mutants are more sensitive to genetic perturbations in the dysregulated pathways, and the reason is likely that they had lower enzyme levels. Thus, removing allosteric feedback inhibition in the arginine, histidine and tryptophan biosynthesis pathways renders these pathways more sensitive against perturbations of gene expression, which may arise in nature due to the stochasticity of gene expression.

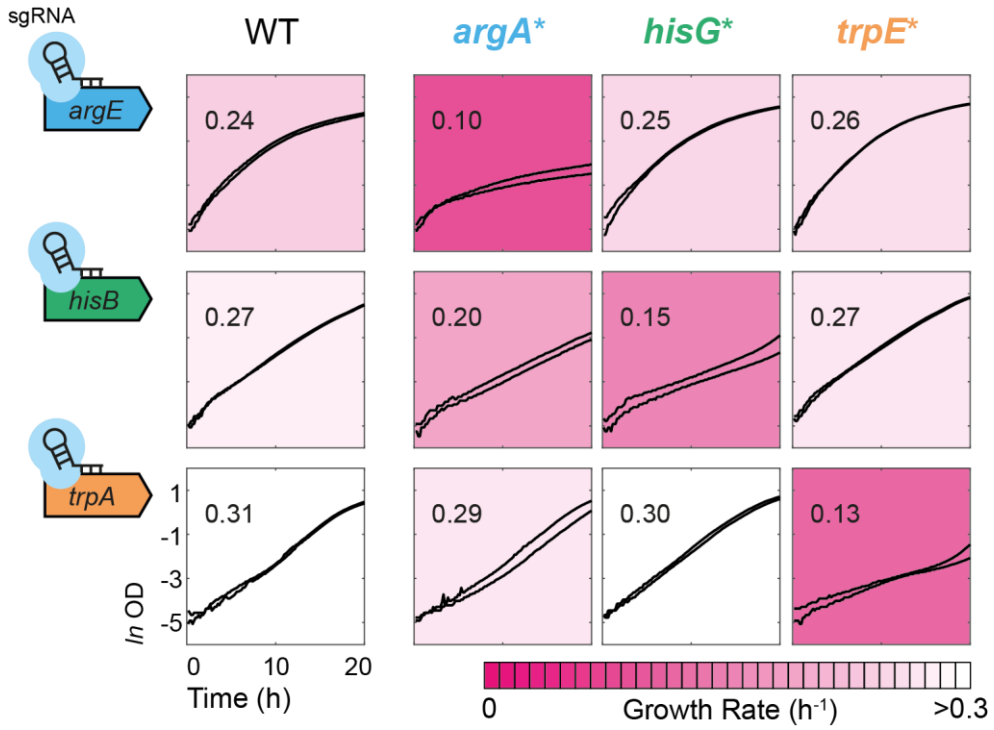


***Figure 19. Enzyme overabundance achieves robustness against perturbations of gene expression by CRISPR interference**. Growth of wild-type (WT),* argA*, hisG*, and* trpE* *with sgRNAs targeting* argE, hisB, *and* trpA. *dCas9 expression was induced with 100 µM IPTG. Growth curves are means of n=3 cultures; two curves per graph show experiments that were performed at different days. Numbers and colors indicate specific growth rates (in h^{-1}), which were estimated by linear regression between 5 and 15 hr. All axes have ranges shown in the lower left graph. This figure is published in Cell Systems 2019, 8 (1), 66-75. e8.*

2.8 Discussion

In this chapter, we studied the interplay between allosteric and transcriptional feedback regulation in amino acid biosynthesis with a mechanistic mathematical model, and multi-omics data of feedback dysregulated *E. coli* mutants. The mechanistic model predicted that allosteric feedback causes higher enzyme levels, and that transcriptional feedback causes lower enzyme levels, while maintaining the same metabolic flux. Experimental data of the dysregulated arginine, histidine and tryptophan strains matched simulated metabolome and proteome changes. This indicated that enzyme levels in amino acid biosynthesis are overabundant, and that the degree of overabundance is balanced by the interplay between allosteric and transcriptional feedback regulation. These results agree with earlier indications that enzymes rarely work at maximum capacity[16,19,20]. Moreover, the model showed that both feedback mechanisms are required to optimize enzyme abundance for two conflicting objectives: robustness and efficiency. To achieve an optimal balance between both objectives, and

to minimize oscillations in amino acid levels, both feedback mechanisms must be equally strong.

A genetic knockdown of metabolic enzymes in the arginine, histidine, and tryptophan pathways by CRISPRi delivered further evidence that allosteric feedback inhibition provides robustness against perturbations of gene expression. While such robustness effects were attributed to allosteric feedback by previous modelling approaches[9,22], we verified our model with *in vivo* data using feedback dysregulated *E. coli* mutants.

Here, we showed the existence of a trade-off between robustness and efficiency in amino acid biosynthesis that emerges from the interplay between allosteric feedback inhibition and transcriptional feedback inhibition. We demonstrated that regulation of enzyme activity and enzyme abundance are not isolated from each other but act in concert to control arginine, histidine, and tryptophan biosynthesis. Together, allosteric feedback and transcriptional feedback set amino acid concentrations, which are signals for enzyme level regulation. Considering that both feedback mechanisms are abundant in metabolic networks, it seems likely that the proposed regulatory principle goes beyond *E. coli* amino acid metabolism[12]. Thus, mathematical models that combine both layers are required to study the interplay between metabolism and enzyme-level regulation in other pathways, or even at a genome scale.

2.9 Methods

Mechanistic model

The stoichiometry of the mechanistic model is shown in **Figure 13**. Mass balancing results in the system of ordinary differential equations (ODEs), F, that is a temporal function of the state variables x (m1, m2, e1, e2) and the kinetic parameters p (k_{cat1}, k_{cat2}, $\beta_{1,max}$, $\beta_{2,max}$, K_1, K_2, K_m, α):

$$F(x,p) = \frac{dx}{dt} = \begin{cases} \frac{dm_1}{dt} = r_1 - r_2 \\ \frac{dm_2}{dt} = r_2 - \alpha\ \mu \\ \frac{de_1}{dt} = \beta_1 - e_1\ \mu \\ \frac{de_2}{dt} = \beta_2 - e_2\ \mu \end{cases} \quad (2.1)$$

The five reactions (r_1, r_2, β_1, β_2, μ) are described by the following kinetic equations:

Reaction 1 is feedback inhibited by m_2 according to normal inhibition kinetics:

$$r_1 = k_{cat,1}\ e_1\ \frac{K_1}{K_1 + m_2} \quad (2.2)$$

In the model without allosteric regulation the equation reduces to:

$$r_1 = k_{cat,1}\ e_1 \quad (2.3)$$

Reaction 2 follows Michaelis-Menten kinetics:

$$r_2 = k_{cat,2}\ e_2\ \frac{m_1}{m_1 + K_m} \quad (2.4)$$

Expression rates of enzyme 1 and enzyme 2 follow inhibition kinetics

$$\beta_1 = \beta_{1,\max}\ \frac{K_2}{K_2 + m_2} \quad (2.5)$$

$$\beta_2 = \beta_{2,\max}\ \frac{K_2}{K_2 + m_2} \quad (2.6)$$

In the model without transcriptional feedback regulation the equations reduce to:

$$\beta_1 = \beta_{1,max} \quad (2.7)$$

$$\beta_2 = \beta_{2,max} \tag{2.8}$$

The growth rate depends on availability of the amino acid:

$$\mu = \mu_{max} \frac{m_2}{m_2 + K_\mu} \tag{2.9}$$

Together, the kinetic equations include eight kinetic parameters k_{cat1}, k_{cat2}, $\beta_{1,max}$, $\beta_{2,max}$, K1, K2, Km and α. The physiological ranges for these parameters were derived from literature values. The boundaries of enzyme turnover number ($k_{cat,1}$ and $k_{cat,2}$) are based on in vitro measured k_{cat} values of enzymes in amino acid biosynthesis (**Appendix,** Error! R eference source not found.) and have values between 930 min^{-1} and 4140 min^{-1}. The maximal enzyme expression rates ($\beta_{1,max}$ and $\beta_{2,max}$) are defined by the translation rate of ribosomes according to **Equation 2.10**. The equation considers the following parameters that were derived from the Bionumbers Database[23]: average translation rate (r_T = 8.4 amino acids s^{-1}), the median and abundance weighted protein length (L = 209 amino acids), the fraction of active ribosomes (f_R = 0.8), the cellular volume ($V_{c,0.6}$ = 3 x 10^{-15} L) and mass ($m_{c,0.6}$ = 9.5 x 10^{-16} kg) at a growth rate of μ = 0.6 h^{-1}, the Avogadro number (N_A = 6.02 x 10^{23} mol^{-1}), the amount of ribosomes per cell at that growth rate ($R_{0.6}$ = 8,000 ribosomes $cell^{-1}$) and the fraction of ribosomes (p) that synthesize the enzyme:

$$\beta_{k,max} = \frac{r_t \cdot R_{0.6} \cdot f_R}{L \cdot N_A \cdot V_c} \cdot p \tag{2.10}$$

The limits of $\beta_{k,max}$ are then derived by varying the fraction of ribosomes (p) that synthesize the enzymes in the pathway. According to the literature, the maximal number for a single enzyme is 7 %[24], therefore we set the boundaries to 1 % and 10 % (p = 0.01 - 0.1). The parameter limits for the Ki and Km values were set to 0.01 mM and 1 mM. The amino acid requirement (α = 86.6 mM) was a fixed parameter based on the average amino acid requirement of an *E. coli* cell (**Appendix, Table 3**). We assumed that the amino acid limits the growth rate reaction only at very low concentrations. This reflects the low K_m values of tRNA ligases. Therefor we fixed Kμ at a low value of 10^{-5} mM and set μ_{max} to the measured growth rate on glucose of 0.6 h^{-1}.

Steady State and Robustness Analysis

For steady state analysis, each parameter set was randomly sampled from a log-uniform distribution within the respective parameter intervals shown above. For each parameter set, the steady state concentrations of e1, e2, m1 and m2 were calculated numerically for each of the three models (complete model, only-transcriptional feedback model and only-allosteric feedback model). Starting values of the numerical solver were 0.01 mM for m1 and m2, and 10^{-5} mM for e1 and e2. The convergence criterion was defined as $<10^{-8}$ change in all variables. To test stability of the steady state we calculated eigenvalues of the Jacobian matrix and tested if all eigenvalues are negative ($\lambda < -10^{-5}$). Unless all three models satisfied the stability and convergence criteria, the parameter set was discarded. This procedure was repeated until 5,000 steady states (with different parameter sets) were achieved. Note that all models share the same parameter sets and reach the same steady state flux.

In order to estimate robustness of the model against perturbations of the maximal enzyme expression rate $\beta_{2,max}$, we used a numerical parameter continuation method. The method is based on finding a connected path of steady state concentrations (x_{SS}: steady state concentration vector containing $e_{1,ss}$, $e_{2,ss}$, $m_{1,ss}$, $m_{2,ss}$), as a parameter, p, is varied. At the beginning of the analysis our system is in a temporal steady state:

$$\frac{dx}{dt} = F(x_{SS}, p) = 0 \tag{2.11}$$

It follows, that the total derivative of $F(x_{SS}, p)$ with respect to the parameters is also zero:

$$\frac{dF(x_{SS}, p)}{dp} = \frac{\delta F}{\delta x_{SS}} \cdot \frac{dx_{SS}}{dp} + \frac{\delta F}{\delta p} = 0 \tag{2.12}$$

After rearranging **Equation 2.12, Equation 2.13** is obtained:

$$\frac{dx_{SS}}{dp} = -\left(\frac{\delta F}{\delta x_{SS}}\right)^{-1} \cdot \frac{\delta F}{\delta p} \tag{2.13}$$

which describes the changes in the steady-state concentrations as a kinetic parameter is varied iteratively. The iteration stops when one of the following two stability criteria is no longer fulfilled. 1st criterion: all real parts of the eigenvalues of the system's Jacobian need to be negative. This implies stability of a steady state. Furthermore, in

Equation 2.13 the inverse of the Jacobian matrix is required. The inversion is only possible if the matrix is regular. Once an eigenvalue reaches zero, the Jacobian becomes singular and matrix inversion is no longer possible. This bifurcation point defines the boundary between the stable and unstable parameter space. In other words: after this point is passed, the system can no longer return to a stable steady state. By checking the eigenvalues of the Jacobian at each step, we make sure that the iteration is terminated when one eigenvalue becomes bigger than $\lambda = -10^{-5}$. 2nd criterion: all variables are required to be positive. The maximum theoretical enzyme amount in the model was calculated as:

$$0 = \beta_{i,max} - e_{i,max}\ \mu \tag{2.14}$$

After rearranging **Equation 2.14** and substituting the upper parameter bound of the maximum protein translation rate ($\beta_{i,max}^{ub}$) an expression for the maximum theoretical enzyme amount of each enzyme is obtained in **Equation 2.15**.

$$e_{i,max} = \frac{\beta_{i,max}^{ub}}{\mu} = \frac{8.5 \cdot 10^{-4}\ mM\ min^{-1}}{0.01\ min^{-1}} = 0.085\ mM \tag{2.15}$$

Considering that the model includes two enzymes, the maximum amount of total enzyme is 0.17 mM, which was defined as the maximal enzyme level (100%).

2.10 References

(1) Chubukov, V.; Gerosa, L.; Kochanowski, K.; Sauer, U. Coordination of Microbial Metabolism. *Nat Rev Microbiol* **2014**, *12* (5), 327–340. https://doi.org/10.1038/nrmicro3238.

(2) Chubukov, V.; Uhr, M.; Le Chat, L.; Kleijn, R. J.; Jules, M.; Link, H.; Aymerich, S.; Stelling, J.; Sauer, U. Transcriptional Regulation Is Insufficient to Explain Substrate-Induced Flux Changes in Bacillus Subtilis. *Mol Syst Biol* **2013**, *9*, 709. https://doi.org/10.1038/msb.2013.66.

(3) ter Kuile, B. H.; Westerhoff, H. V. Transcriptome Meets Metabolome: Hierarchical and Metabolic Regulation of the Glycolytic Pathway. *FEBS Lett* **2001**, *500* (3), 169–171. https://doi.org/10.1016/s0014-5793(01)02613-8.

(4) Reznik, E.; Christodoulou, D.; Goldford, J. E.; Briars, E.; Sauer, U.; Segrè, D.; Noor, E. Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Rep* **2017**, *20* (11), 2666–2677. https://doi.org/10.1016/j.celrep.2017.08.066.

(5) Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Res* **2002**, *30* (1), 47–49. https://doi.org/10.1093/nar/30.1.47.

(6) Hirasawa, T.; Shimizu, H. Recent Advances in Amino Acid Production by Microbial Cells. *Current Opinion in Biotechnology* **2016**, *42*, 133–146. https://doi.org/10.1016/j.copbio.2016.04.017.
(7) Reaves, M. L.; Young, B. D.; Hosios, A. M.; Xu, Y.-F.; Rabinowitz, J. D. Pyrimidine Homeostasis Is Accomplished by Directed Overflow Metabolism. *Nature* **2013**, *500* (7461), 237–241. https://doi.org/10.1038/nature12445.
(8) Hofmeyr, J.-H. S.; Cornish-Bowden, A. Regulating the Cellular Economy of Supply and Demand. *FEBS Letters* **2000**, *476* (1), 47–51. https://doi.org/10.1016/S0014-5793(00)01668-9.
(9) Grimbs, S.; Selbig, J.; Bulik, S.; Holzhütter, H.-G.; Steuer, R. The Stability and Robustness of Metabolic States: Identifying Stabilizing Sites in Metabolic Networks. *Mol Syst Biol* **2007**, *3*, 146. https://doi.org/10.1038/msb4100186.
(10) Kacser, H.; Burns, J. A. The Control of Flux. *Biochem Soc Trans* **1995**, *23* (2), 341–366. https://doi.org/10.1042/bst0230341.
(11) Goyal, S.; Yuan, J.; Chen, T.; Rabinowitz, J. D.; Wingreen, N. S. Achieving Optimal Growth through Product Feedback Inhibition in Metabolism. *PLOS Computational Biology* **2010**, *6* (6), e1000802. https://doi.org/10.1371/journal.pcbi.1000802.
(12) Sander, T.; Farke, N.; Diehl, C.; Kuntz, M.; Glatter, T.; Link, H. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. Coli by Enforcing Enzyme Overabundance. *Cell Syst* **2019**, *8* (1), 66-75.e8. https://doi.org/10.1016/j.cels.2018.12.005.
(13) Schmidt, A.; Kochanowski, K.; Vedelaar, S.; Ahrné, E.; Volkmer, B.; Callipo, L.; Knoops, K.; Bauer, M.; Aebersold, R.; Heinemann, M. The Quantitative and Condition-Dependent Escherichia Coli Proteome. *Nat Biotechnol* **2016**, *34* (1), 104–110. https://doi.org/10.1038/nbt.3418.
(14) Zaslaver, A.; Mayo, A. E.; Rosenberg, R.; Bashkin, P.; Sberro, H.; Tsalyuk, M.; Surette, M. G.; Alon, U. Just-in-Time Transcription Program in Metabolic Pathways. *Nat Genet* **2004**, *36* (5), 486–491. https://doi.org/10.1038/ng1348.
(15) O'Brien, E. J.; Utrilla, J.; Palsson, B. O. Quantification and Classification of E. Coli Proteome Utilization and Unused Protein Costs across Environments. *PLOS Computational Biology* **2016**, *12* (6), e1004998. https://doi.org/10.1371/journal.pcbi.1004998.
(16) Davidi, D.; Milo, R. Lessons on Enzyme Kinetics from Quantitative Proteomics. *Current Opinion in Biotechnology* **2017**, *46*, 81–89. https://doi.org/10.1016/j.copbio.2017.02.007.
(17) Kotte, O.; Zaugg, J. B.; Heinemann, M. Bacterial Adaptation through Distributed Sensing of Metabolic Fluxes. *Mol Syst Biol* **2010**, *6*, 355. https://doi.org/10.1038/msb.2010.10.
(18) Savageau, M. A.; Voit, E. O. Recasting Nonlinear Differential Equations as S-Systems: A Canonical Nonlinear Form. *Mathematical Biosciences* **1987**, *87* (1), 83–115. https://doi.org/10.1016/0025-5564(87)90035-6.
(19) Hackett, S. R.; Zanotelli, V. R. T.; Xu, W.; Goya, J.; Park, J. O.; Perlman, D. H.; Gibney, P. A.; Botstein, D.; Storey, J. D.; Rabinowitz, J. D. Systems-Level Analysis of Mechanisms Regulating Yeast Metabolic Flux. *Science* **2016**, *354* (6311), aaf2786. https://doi.org/10.1126/science.aaf2786.
(20) Fendt, S.-M.; Buescher, J. M.; Rudroff, F.; Picotti, P.; Zamboni, N.; Sauer, U. Tradeoff between Enzyme and Metabolite Efficiency Maintains Metabolic Homeostasis

upon Perturbations in Enzyme Capacity. *Mol Syst Biol* **2010**, *6*, 356. https://doi.org/10.1038/msb.2010.11.

(21) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metabolic Engineering* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.

(22) Chandra, F. A.; Buzi, G.; Doyle, J. C. Glycolytic Oscillations and Limits on Robust Efficiency. *Science* **2011**, *333* (6039), 187–192. https://doi.org/10.1126/science.1200705.

(23) Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers—the Database of Key Numbers in Molecular and Cell Biology. *Nucleic Acids Res* **2010**, *38* (Database issue), D750–D753. https://doi.org/10.1093/nar/gkp889.

(24) Li, G.-W.; Burkhardt, D.; Gross, C.; Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **2014**, *157* (3), 624–635. https://doi.org/10.1016/j.cell.2014.02.033.

Chapter 3 A mathematical model predicts robust regulation in engineered glycerol-producing *E. coli*

Wang, C.-Y.[#]; Lempp[#], M.; Farke, N.; Donati, S.; Glatter, T.; Link, H. Metabolome and Proteome Analyses Reveal Transcriptional Misregulation in Glycolysis of Engineered E. Coli. Nat Commun 2021, 12, 4929. https://doi.org/10.1038/s41467-021-25142-0.
[#]Authors contributed equally

Parts of this chapter are published in Nature Communications 2021, 12, 4929. My contribution to this work included the design of the mechanistic mathematical models, and the analysis of the mechanistic mathematical models.

3.1 Abstract

Synthetic metabolic pathways are a burden for engineered bacteria, but the underlying mechanisms often remain elusive. In this chapter, we developed a mechanistic mathematical model of glycolysis and glycerol pathway that combines metabolism and transcriptional regulation to study metabolic burden in engineered glycerol-producing *E. coli*. The model showed that transcriptional regulation of glycolysis causes system failure when flux is drained from glycolysis to produce glycerol. Draining flux from glycolysis for glycerol production depletes fructose-1,6-bisphosphate (FBP) levels. Low FBP levels activate Cra, causing downregulation of glycolytic enzymes and upregulation of gluconeogenesis enzymes, impairing *E. coli*'s ability to grow on glucose. The model predicted that engineering Cra regulation into the glycerol pathway leads to a more robust glycerol production strain, solving the regulatory problem. Finally, we verified the model predictions by experimentally creating a robust glycerol-producing *E. coli* strain.

3.2 Introduction

Expanding the metabolic network with synthetic metabolic pathways enables the production of valuable metabolites or allows cells to gain access to new feedstocks like CO_2[1,2]. However, modifying or extending endogenous metabolism can negatively affect fitness and cellular growth of the host. For example, overproduction of certain chemicals depletes metabolites which are then no longer available for growth. This competition between endogenous and synthetic metabolism causes metabolic burden that leads to stress responses and physiological changes of the host[3]. Therefore, a current challenge is to minimize metabolic burden, while optimizing resource allocation, i.e. metabolic flux, through the synthetic pathways.

Most theoretical approaches that are concerned with optimizing yield and productivity of microbes use flux balance analysis (FBA)[2,4], or similar stoichiometric approaches[5]. However, these approaches don't consider how synthetic metabolic pathways interfere with a native network that has optimized its structure and kinetic parameters for growth[6,7]. Hence, enzyme levels or kinetic parameters of the native system may simply not tolerate such interferences. To assess the viability of engineered pathways,

theoretical approaches use sensitivity analyses[8] and robustness analyses[9] to test if metabolic pathways and their kinetic parameters are robust[6,9,10]. For example, the ensemble modelling for robustness analysis (EMRA)[9] method uses a numerical continuation method to perturb parameters in metabolic models until instabilities occur. This method was used to assess the robustness of native and non-native metabolic pathways,[7,11] as well as synthetic gene-metabolic oscillators[12].

To achieve viable and robust synthetic pathways, dynamic control of the enzyme expression levels needs to be engineered. While models have shown the benefits of adding synthetic regulation to pathways[10], engineering allosteric regulation experimentally is a challenging task, because protein structures need to be altered[13]. Conversely, it is more feasible to engineer control of gene expression in synthetic pathways. This is commonly done by engineering protomers to allow binding of certain transcription factors (TFs). Conventionally, the TF activity is then controlled by metabolites of the synthetic pathway. The resulting interplay between gene expression and metabolism has improved production rates of lycopene[14], fatty acids[15], and precursors of isoprenoids[16].

Here, we developed a mechanistic mathematical model of glycolysis that combines metabolism and transcriptional regulation to study metabolic burden in an *E. coli* strain that was engineered to overproduce glycerol. Induction of the synthetic glycerol pathway caused growth burden and lower glycerol titers. The mechanistic model predicted that transcriptional misregulation by Cra is the reason for this behaviour: Induction of the glycerol pathway depletes FBP, which activates Cra. Subsequently, Cra inhibits glycolysis and activates gluconeogenesis, causing lower growth rates and lower glycerol titers. To solve the regulatory problem, the mechanistic model predicted that controlling the glycerol pathway with Cra increases robustness of the production strain. Finally, we engineered Cra regulation in the glycerol pathway. This way, we validated the model predictions by achieving higher growth rates and higher glycerol production rates. Thus, a mechanistic model that combines metabolism and transcriptional regulation, enabled the engineering of a robust glycerol production strain.

3.3 Glycerol overproduction in *E. coli* causes growth burden

To study how a synthetic pathway impacts metabolism, we engineered the glycerol biosynthesis pathway from *Saccharomyces cerevisiae* (*S. cerevisiae*) in *E. coli* (**Figure 20a**). The synthetic glycerol pathway takes dihydroxyacetone phosphate (DHAP) and converts it via the enzyme glycerol-3-phosphate dehydrogenase 1 (GPD1) to glycerol phosphate. In a final step, glycerol phosphate is converted to glycerol via the enzyme glycerol-3-phosphate phosphohyrolase 2 (GPP2). To control glycerol production, we used an arabinose inducible pBAD promoter. We induced the glycerol pathway at different arabinose concentrations and measured growth rate and glycerol titers. However, we already observed growth defects at low arabinose levels (**Figure 20b**). Glycerol titers increased at low arabinose levels but decreased at higher arabinose levels (**Figure 20c**), thus indicating metabolic burden caused by induction of the glycerol pathway.

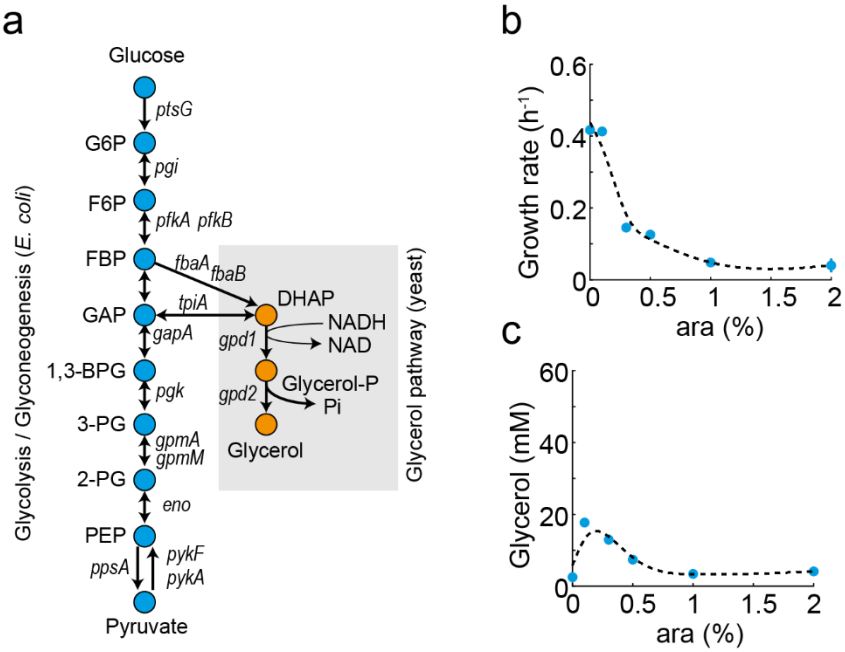


Figure 20. Glycolysis and an engineered glycerol pathway a) *Metabolic network of glycolysis in E. coli and the synthetic glycerol pathway from yeast (grey box). The synthetic pathway consists of two enzymes from S. cerevisiae: glycerol-3-phosphate dehydrogenase 1 (GPD1) and glycerol-3-phosphate phosphohydrolase 2(GPP2). b) Growth was measured in a plate reader at different induction levels of GPD1 (0, 0.1, 0.3, 0.5, 1, and 2% ara). Glycerol in the medium was measured after 24 h. Grwoth rates were determined by regression analysis between 5 and 10 h. Growth curves and dots show the means of n = 2 plate reader cultures. Parts of this figure are published in Nature Communications* ***2021****, 12, 4929.*

We then sought to understand the mechanisms that caused the growth burden. One possible reason for the burden is the competition between glycolysis flux and glycerol flux. Flux that is channelled through the glycerol pathway is no longer available for growth. To test this, we used flux balance analysis (FBA) with a genome-scale model of *E. coli* and showed that growth rate and glycerol flux follow a theoretical trade-off. To test if the strain followed this theoretical trade-off, we measured glycerol production rates and growth rates at three induction levels (arabinose): 0, 0.1, and 0.5 % arabinose (**Figure 21**). However, the experimental trade-off and the theoretical trade-off did not match. At 0.5 % arabinose, the measured growth rates were below the theoretical line, indicating that flux alone is insufficient to explain the phenotypes. Therefore, we next developed a small mechanistic model of glycolysis and glycerol production that includes transcriptional regulation (enzyme-level regulation).

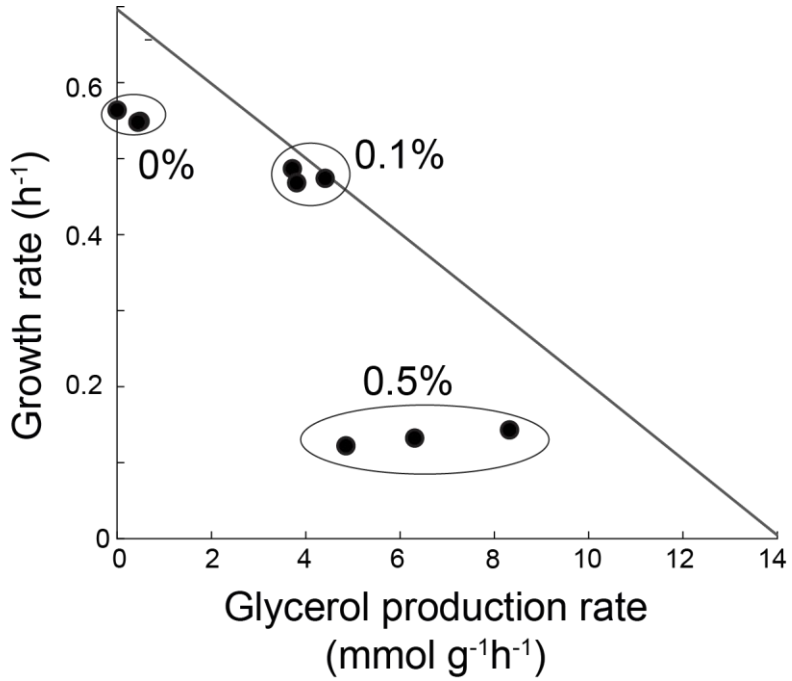


Figure 21. Theoretical and experimental relationships between the glycerol production rate and the growth rate of E. coli*. The line is achieved via flux balance analysis with a genome-scale model of E. coli metabolism (iML1515). Dots are growth rates and glycerol production rates measured in shake flask cultures of the base strain at 0, 0.1, and 0.5% arabinose. The figure is published in Na*ture *Commun*ications *2021, 12, 4929.*

3.4 A mechanistic model predicts robust transcriptional regulation for glycerol producing *E. coli*

To better understand the mechanisms that cause the growth defects, we developed a small mechanistic model that combines metabolism and enzyme-level regulation (**Figure 22**). The model included one metabolite (FBP) and two enzymes e1 and e2. Enzyme e1 corresponds to glyceraldehyde-3-phosphate dehydrogenase (GapA) in lower glycolysis and e2 is GPD1 in the glycerol pathway. FBP is the central metabolite that is consumed by lower glycolysis ($r_{lower\_glycolysis}$) and by the glycerol pathway ($r_{glycerol}$). Both reactions follow Michaelis-Menten kinetics, which are a well-established kinetic format for enzymatic reactions[17]. Like flux balance analysis (**Figure 21**), we assumed constant influx in the model and fixed the reaction rate in upper glycolysis to 4.9 mmol g^{-1} h^{-1}. This means that FBP is produced at a constant rate and is either used for glycerol production, or for growth according to the mass balance in **Equation 3.1.**

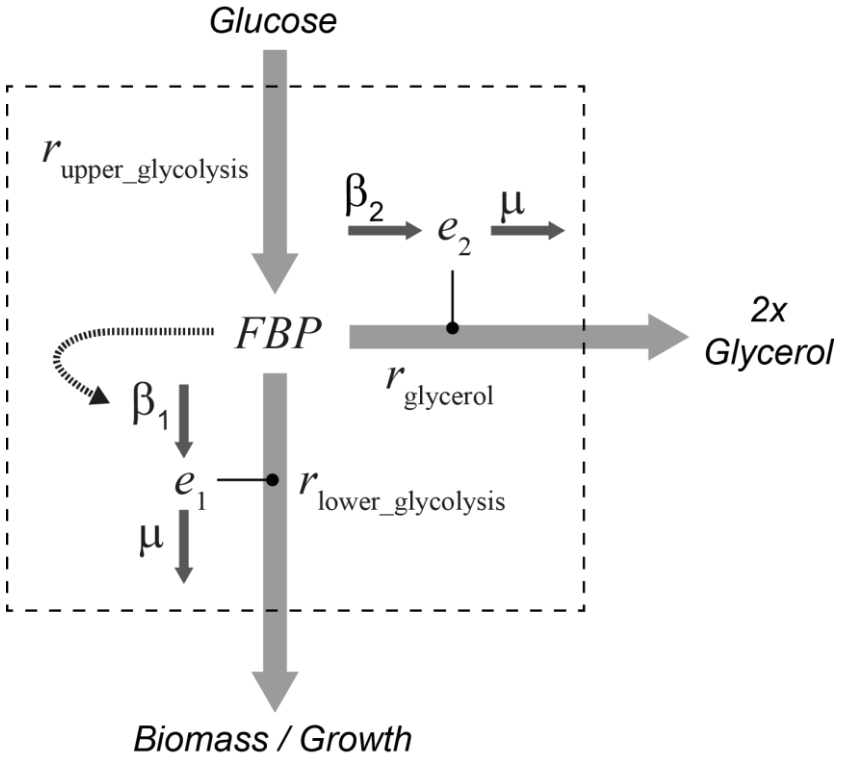


***Figure 22**. **Stoichiometry and structure of the kinetic model**. The dashed box is the model boundary. FBP is a metabolite in glycolysis, and e1 and e2 are enzymes. Solid, gray arrows are metabolic reactions. The input reaction in upper glycolysis is fixed to 4.9 mmol g^{-1} h^{-1}. The reaction in lower glycolysis and the glycerol pathway depends on FBP levels and enzyme levels, according to Michaelis–Menten kinetics. The dashed arrow is FBP-activation of β_1 (the expression rate of e1) and represents the net effect of Cra regulation: FBP inhibits Cra; Cra inhibits β_1. Gradual increases of β_2 (the expression rate of e2) simulate induction with arabinose. The figure is published in Nature Communications 2021, 12, 4929*

In total, we analysed three models with different feedback regulation (**Figure 23a**). In the base model, FBP only activates the expression of enzyme e1. This interaction resembled transcriptional regulation of lower glycolysis by Cra. The second model is the Δcra model of the ΔCra strain, which has no feedback regulation. In a third model (2x Cra model), expression of both enzymes e1 and e2 is feedback regulated by FBP levels. To simulate Cra regulation, we used a power-law term that affects the maximal enzyme production rate. Since the power-law term equals one in the un-induced state, all models share the same parameter set. Consequently, all models are at the same steady state at the beginning of the simulations, which allows for an easier comparison of the simulation results.

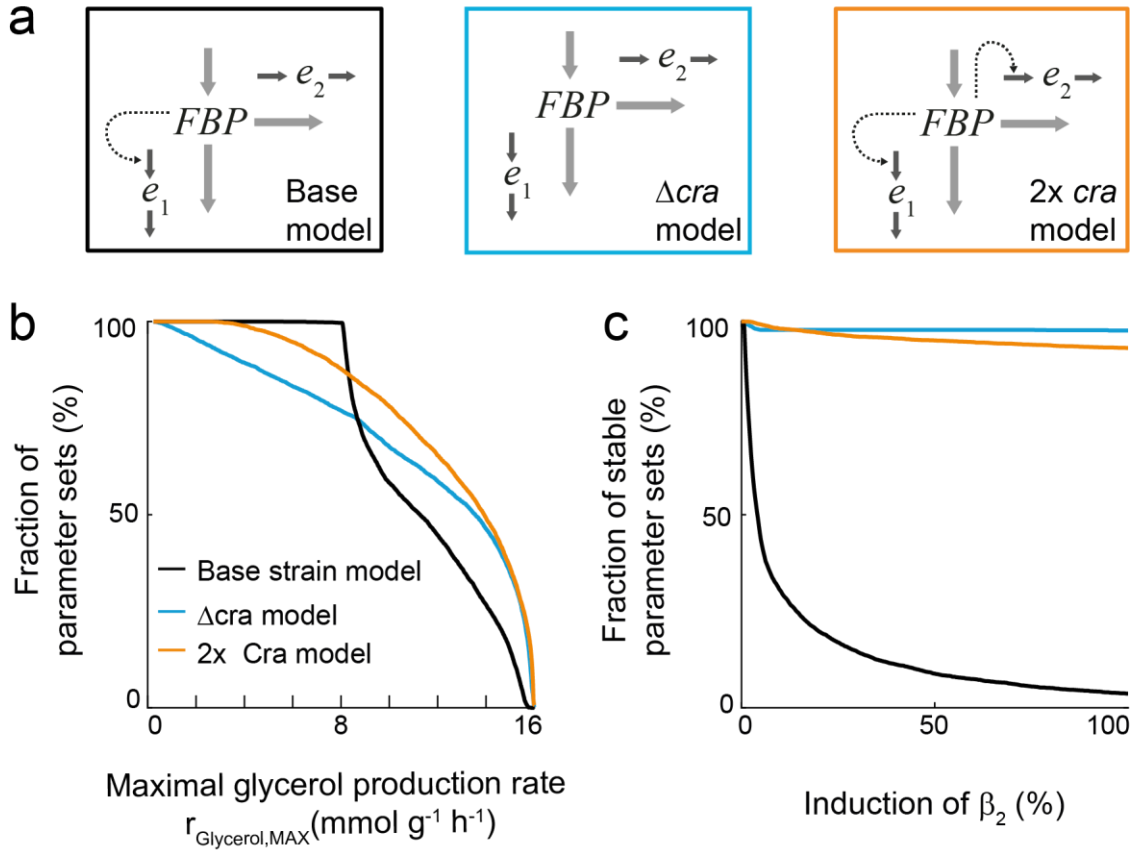


Figure 23. Robustness analysis of three models. a) *A model of the base strain (black), a model of the Δcra strain (blue), and a model with additional Cra regulation of e2 expression (orange).* ***b)*** *The three models in b were simulated with the same 5,000 parameter sets that were obtained by random sampling (**Table 1**). For each parameter set, β2 (the synthesis rate of e2) was increased until the model became unstable or until the expression rate of β2 reached the maximum. Shown is the maximal glycerol flux that was achieved with each model as a cumulative sum distribution.* ***c)*** *Robustness is shown as the percentage of the 5,000 models that remain stable at a given induction level. This figure is published in Nature Communications 2021, 12, 4929.*

The three models were analysed with 5,000 parameter sets that were randomly sampled from physiologically meaningful ranges based on literature values (**Table 1**). We sampled the power-law exponent between 1 and 2, to ensure that Cra-regulation depends at least linearly on the concentration of FBP, and to avoid instabilities that can occur at exponents >2. For each of the 5,000 parameter sets, we calculated the maximal glycerol production rate ($r_{glycerol,max}$) that can possibly be achieved given the specific set of parameters. To estimate $r_{glycerol,max}$, we employed a numerical continuation method[9], which iteratively increases the expression rate of enzyme e2 ($\beta 2$) and computes the new steady state for FBP, e1, and e2. After each iteration, the continuation method determines the stability of the model by inspecting the eigenvalues of the Jacobian matrix and terminates if instabilities occur in the model. If the model remains stable, the continuation method terminates at the maximal expression rate of e2 ($\beta_{2,max}$), which we defined as the rate in which 20 % of the ribosomes translate e2. Thus, $r_{glycerol,max}$ is the glycerol production rate at the termination point of the continuation method and we obtained 5,000 values of $r_{glycerol,max}$ for each of the three models.

The distribution of the 5,000 $r_{glycerol,MAX}$ values showed that the Δ*cra* model performed better than the base strain model, because more parameter sets achieved higher maximal glycerol production rates ($r_{glycerol,MAX}$) with the Δ*cra* model than with the base strain model (**Figure 23b**). The underlying assumption was: the more parameter sets achieve high glycerol fluxes, the higher the likelihood that the real system would achieve them too. The model of the base strain did not achieve high glycerol production rates, because the model was not stable at higher induction levels (**Figure 23c**).

To better understand the origin of these instabilities, we performed time-course simulations with the three models using an average parameter set (**Figure 24**, **Table 1**). The time-course simulations matched the results obtained with the continuation method, thus confirming that both numerical approaches yield the same results. We simulated the models at different induction levels.

The base model was not stable at higher induction, because enzyme e2 increased exponentially despite a linear increase of $\beta 2$. Thus, there is a critical point at which the expression rate of e2 exceeds its dilution by growth. These imbalances are probably amplified by Cra-regulation, since Cra downregulates *e1* and therefore also the growth

rate. In contrast, the Δ*cra* model was stable at almost all induction levels. The best model in our analysis was the 2x*cra* model. With this model, the highest fraction of parameter sets achieved high glycerol fluxes (**Figure 23b**).

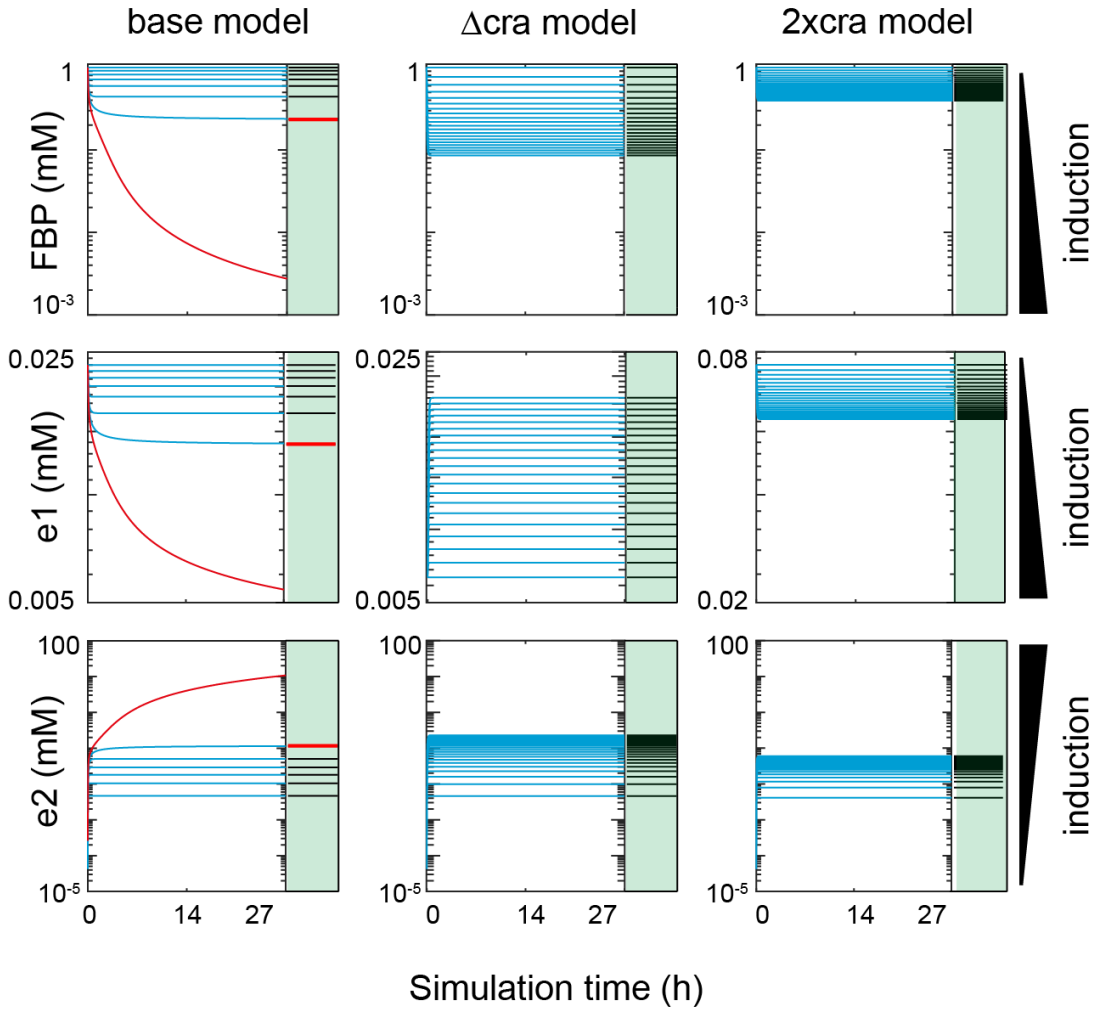


Figure 24. Comparison time-course simulation and continuation method*. Time-course simulation of the three models with an average parameter set and results with the parameter continuation method (green panel). Each of the three models (base model, 2xCra model, Δcra model) was simulated with the average parameter set (**Table 1**). Blue lines are different induction levels ranging from ind = 0 until 1 (for the 2xcra model,Δcra model), or until the calculated bifurcation point (for the base strain model). The red line for the base model is the time-course of an induction level that is 1% stronger than at the bifurcation point. The panels with green background show the steady states calculated by the continuation method. Black lines are stable steady states and a red line indicates a bifurcation point. The figure is published in Nature Communications 2021, 12, 4929*

Moreover, the robustness of the 2x*cra* model was comparable to the robustness of the Δ*cra* model (**Figure 23c**). In conclusion, the model predicted that engineering Cra-regulation into the glycerol pathway should lead to higher glycerol production rates, and we next tested this prediction experimentally.

3.5 A Cra-regulated pBAD promoter improves growth rate and glycerol titers

Next, we sought to verify the model predictions experimentally. Therefore, we put the pBAD promoter under the control of Cra (pBAD-Cra strain) and created a Cra knock-out strain (Δcra strain). As predicted by the model, the pBAD-Cra strain had the highest growth rate of all three strains and even maintained growth at full induction with 2 % arabinose (**Figure 25a**). Moreover, the pBAD-Cra strain also achieved the highest glycerol titers at induction with 1 % arabinose (**Figure 25b**). In contrast, the Δcra strain could not tolerate high ara levels and achieved less than half glycerol titers compared to the pBAD-Cra strain at induction with 1 % arabinose. These data confirm the model prediction that a doubly Cra-regulated strain performs better than the base strain and the Δcra strain.

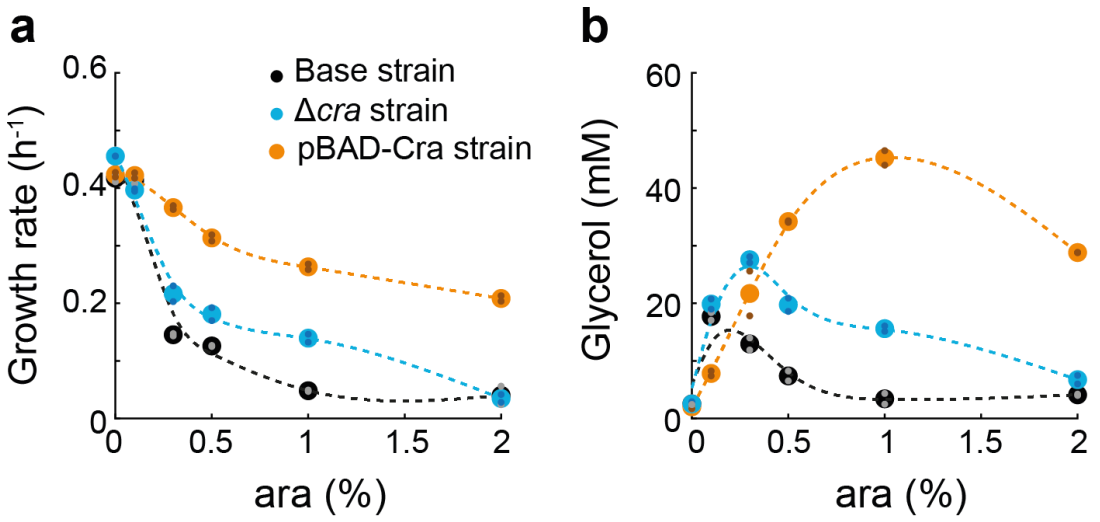


Figure 25. Growth rates and glycerol titers. ***a)*** *Growth rate.* ***b)*** *Glycerol titers. The pBAD-Cra strain (orange) was cultured in 96-well plates. Growth was measured in a plate reader at different induction levels of GPD1 (0, 0.1, 0.3, 0.5, 1, and 2% ara). Glycerol in the medium was measured after 24 h. Growth rates were determined by regression analysis between 5 and 10 h. Small dots show data from* n *= 2 plate reader cultures and big dots are the mean. Data of the base strain (black) and the Δ*cra *strain (blue) are shown as a reference. The figure is published in Nature Communications 2021, 12, 4929.*

3.6 Discussion

Engineering of metabolic pathways often causes metabolic burden. However, the underlying reasons are often not well understood. Here, we developed a mathematical model of glycolysis that combines metabolism and enzyme-level regulation by the transcription factor Cra to study metabolic burden in glycerol-producing *E. coli*. Induction of the glycerol pathway impaired *E. coli's* ability to grow and produce glycerol on glucose minimal medium. The mechanistic model predicted that these effects are caused by transcriptional misregulation. Draining flux from the glycolysis depletes FBP levels, which activates Cra. Cra reduces levels of glycolytic enzyme, amplifying the effect, and subsequently causing exponential changes of metabolite and enzyme levels. This resembles the inactivation of glycolysis and the concomitant activation of gluconeogenesis by Cra, causing slow growth rates and glycerol titers on glucose minimal medium. To solve this regulatory problem, the model suggested to put the glycerol pathway under transcriptional control of Cra. The 2xCra model achieved higher glycerol production rates and tolerated maximum induction of the glycerol pathway. We then engineered Cra regulation into the glycerol pathway (pBAD-Cra strain). The pBAD-Cra strain achieved the highest growth rates and glycerol titers of all three strains, thus verifying the model predictions.

Engineered metabolic pathways often lack regulatory circuits that maintain stable metabolite concentrations. These results emphasize the importance of maintaining regulatory metabolites above a critical threshold in engineered microbes. To identify the right regulatory mechanisms, mathematical models have been used to suggest stabilizing sites[10,14], but also to understand the causes of these instabilities[9]. Commonly production strains in biotechnology are optimized with stoichiometric models[4,5]. Recent years, however, have seen a rise of kinetic models that use regulatory information and robustness analyses to optimize performance of production strains[18–21]. Yet, none of these approaches consider the interplay between metabolism and transcriptional regulation, and how this interaction affects metabolic burden.

Here, we developed a small mechanistic model of the branch point between glycolysis and glycerol pathway that combines metabolism and transcriptional regulation. The

model revealed that transcriptional regulation of the base model is sub-optimal for glycerol production and that engineered Cra regulation of the glycerol pathway maintains regulatory metabolites above critical thresholds This supports the hypothesis that Cra-dependent regulation counteracts a decline in the concentration of FBP by downregulating the expression of the glycerol pathway in response to decreasing FBP levels. However, as it remains an open question whether this regulation is truly dynamic, we cannot rule out the possibility that, due to the constant inhibitory activity of Cra, the pBAD-Cra promoter simply functions as a weaker pBAD promoter. Future studies should clarify whether the pBAD-Cra promoter automatically adapts to new conditions, e.g., by shifting the glycerol producers between different environments.

3.7 Methods

Mechanistic modelling and steady-state analysis

The stoichiometry of the model is shown in **Figure 22**. Mass balancing yields a system of ordinary differential equations (ODEs), F, that is a temporal function of the state variables x and the kinetic parameters p:

$$F(x,p)=\frac{dx}{dt}=\begin{cases}\frac{dFBP}{dt}=r_{upper\_glycolysis}-r_{lower\_glycoylsis}-r_{glycerol}-FBP\cdot\mu\\ \frac{de_1}{dt}=\beta_1-e_1\cdot\mu\\ \frac{de_2}{dt}=\beta_2-e_2\cdot\mu\end{cases} \quad (3.1)$$

The metabolite FBP is produced by $r_{upper\_glycolysis}$ and consumed by $r_{lower\_glycolysis}$ and $r_{glycerol}$. Moreover, FBP is diluted by growth. The enzyme e1 is a lower glycolysis enzyme for which we used parameters of glyceraldehyde-3-phosphate dehydrogenase (GapA) and e2 is GPD1. Both enzymes are produced by a production term β and they are removed by dilution by growth. We assumed that enzyme degradation contributes little to the overall enzyme turnover and therefore can be neglected.

An upper glycolytic flux of 4.904 mmol $g^{-1}h^{-1}$ was estimated with FBA using a glucose uptake rate of 8 mmol g^{-1} h^{-1}. With the specific cell volume for *E. coli* (2 µl mg^{-1})[22] the reaction rate $r_{upper\ glycolysis}$ is:

$$r_{upper\_glycolysis}=\frac{4.904\ mmol\ g^{-1}h^{-1}}{0.002\ l\ g^{-1}}*\frac{h}{60\ min}=40.87\frac{mM}{min} \quad (3.2)$$

The reactions $r_{lower\_glycolysis}$ and $r_{glycerol}$ follow Michaelis-Menten kinetics:

$$r_{lower\_glycolysis}=k_{cat,1}\cdot e_1\cdot\frac{FBP}{FBP+K_{m1}} \quad (3.3)$$

$$r_{glycerol}=k_{cat,2}\cdot e_2\cdot\frac{FBP}{FBP+K_{m2}} \quad (3.4)$$

The expression rates of enzyme 1 (GapA) and enzyme 2 (GPD1) are:

$$\beta_1=\beta_{1,max}\cdot\left(\frac{FBP}{FBP_{SS}}\right)^{\alpha1} \quad (3.5)$$

and

$$\beta_2 = \beta_{2,max} \cdot ind \cdot \left(\frac{FBP}{FBP_{SS}}\right)^{\alpha 2} \quad (3.6)$$

Cra-regulation was simulated with a power law term $\left(\frac{FBP}{FBP_{SS}}\right)^{\alpha}$ that affects the maximal enzyme expression rate. The power-law format has the advantage that the power-law term equals one in the un-induced state and therefore allows the same parameter values for the base model, the Δ*cra* model and the 2x *cra* model. Further, setting α to zero removes the regulation. Therefore, $\alpha 2$ was zero in the base model, while $\alpha 1$ and $\alpha 2$ were zero in the Δ*cra* model.

We assumed that the growth rate μ is proportional to $r_{lower\_glycolysis}$, because previous ^{13}C-labelling data showed a positive correlation between lower glycolytic flux and growth in *E. coli*[23]. With a growth rate of 0.01 min^{-1} in the un-induced state, the proportionality factor alpha follows as:

$$r_{lower\ glycolysis} = r_{upper\ glycolysis} - FBP \cdot \mu = 40.86\ mM\ min^{-1} \quad (3.7)$$

$$alpha = \frac{r_{lower\_glycolysis}}{\mu} = \frac{40.86\ mM\ min^{-1}}{0.01\ min^{-1}} = 4086\ mM \quad (3.8)$$

$$\mu = \frac{r_{lower\_glycolysis}}{alpha} \quad (3.9)$$

In total, the model includes eight kinetic parameters $k_{cat,1}$, $k_{cat,2}$, $K_{m,1}$, $K_{m,2}$, $\beta_{1,max}$, $\beta_{2,max}$, $\alpha 1$, and $\alpha 2$. The parameters were either sampled 5000 times from a log-uniform distribution from predefined intervals or calculated based on steady state constraints. $K_{m,1}$ and $K_{m,2}$ were randomly sampled between 0.01 and 10 mM to account for high and low saturation of enzymes. The power-law exponents $\alpha 1$ and $\alpha 2$ were randomly sampled between 1 and 2. The lower bound was 1 to ensure that the expression rate is at least linearly dependent on the FBP concentration. The upper bound was 2 to avoid higher-order dynamics that can cause instabilities[24]. The $k_{cat,2}$ value was based on the kinetic parameter of GPD1 ($k_{cat,2}$ = 1705 min^{-1})[25] and was sampled between 0.33-fold and 3-fold of this literature value. The parameter $k_{cat,1}$ followed from the steady state constraint of the un-induced state where $r_{glycerol}$ = 0. $\beta_{1,max}$ was derived from the mass balances of e_1, assuming steady state:

$$\beta_{1,max} = e_1 \cdot \mu \quad (3.10)$$

The concentration of *e1* was 0.0238 mM, based on quantitative proteome data for GapA[14] resulting in $\beta_{1,max}$ = 0.000238 mM min^{-1}. The maximal enzyme expression rate in the glycerol pathway ($\beta_{2,max}$) was defined by the translation rate of ribosomes according to:

$$\beta_{2,max} = \frac{r_t \cdot R_{0.6} \cdot f_R}{L \cdot N_A \cdot V_c} \cdot p = 0.0017\ mM\ min^{-1} \quad (3.11)$$

Equation 3.11 considers the following parameters that were derived from the Bionumbers Database[26]: average translation rate (r_t = 8.4 amino acids s^{-1}), the median and abundance weighted protein length (L = 209 amino acids), the fraction of active ribosomes (f_R = 0.8) and the cellular volume ($V_{c,0.6}$ = 3 x 10^{-15} L) and at a growth rate of μ = 0.6 h^{-1}, the Avogadro number (N_A = 6.02 x 10^{23} mol^{-1}), the amount of ribosomes per cell at that growth rate ($R_{0.6}$ = 8000 ribosomes $cell^{-1}$). The fraction of ribosomes (*p*) that synthesize GPD1 at full induction was assumed to be 20%, because only 50% of the ribosomes can translate a heterologous protein and this is already associated with significant protein burden[27].

***Table 1**. Values and units of parameters and variables*

Parameter / Species	Lower bound	Upper bound	Unit
$r_{upper\_glycolysis}$	40.87	40.87	mM min^{-1}
$k_{cat,1}$	calculated		min^{-1}
$K_{m,1}$	0.01	10	mM
$k_{cat,2}$	511.5	5115	min^{-1}
$K_{m,2}$	0.01	10	mM
$\beta_{1,max}$	calculated		mM min^{-1}
$\beta_{2,max}$	0.0017	0.0017	mM min^{-1}
ind	0	1	-
α1	1	2	-
α2	1	2	-
$\mu_{initial}$	0.01	0.01	min^{-1}
FBP	1	1	mM
e1(GapA)	0.0238	0.0238	mM
e2(GPD1)	0	0	mM

Steady state and robustness analysis

To obtain steady states of the un-induced system, $\beta_{2,max}$ and *e2* were set to zero. Then six parameters were randomly sampled from intervals defined above. The two parameters (k_{cat1} and $b_{1,max}$) were calculated to ensure steady state conditions. To test stability of the steady states, eigenvalues of the Jacobian matrix were calculated, and tested if all eigenvalues are negative ($\lambda < -10^{-5}$). The procedure was repeated until 5,000 stable steady states were achieved. Next, induction (*ind* in **Equation 3.6**) was iteratively increased from 0 to 1 using a numerical parameter continuation method. The method is based on finding a connected path of steady state concentrations (x_{ss}: steady state concentration vector containing e_{1ss}, e_{2ss}, FBP_{ss}) as a parameter p is varied. As the system is in steady state it follows that:

$$\frac{dx}{dt} = F(x_{ss}, p) = 0 \tag{3.12}$$

The derivative of $F(x_{ss},p)$ with respect to the parameters is also zero:

$$\frac{dF(x_{ss}, p)}{dp} = \frac{\delta F}{\delta x_{ss}} \cdot \frac{dx_{ss}}{dp} + \frac{\delta F}{\delta p} = 0 \tag{3.13}$$

After rearranging **Equation 3.13**, **Equation 3.14** is obtained:

$$\frac{dx_{ss}}{dp} = -\left(\frac{\delta F}{\delta x_{ss}}\right)^{-1} \cdot \frac{\delta F}{\delta p} \tag{3.14}$$

which describes the changes in the steady-state concentrations as a kinetic parameter is varied iteratively. The iteration stops when one of the following two stability criteria is no longer fulfilled. 1st criterion: all real parts of the eigenvalues of the system's Jacobian need to be negative. In **Equation 3.14** the inverse of the Jacobian Matrix ($\delta F/\delta x_{SS}$) is required. The inversion is only possible if the matrix is regular. Once an eigenvalue reaches zero, the Jacobian becomes singular and matrix inversion is no longer possible. This bifurcation point defines the boundary between the stable and unstable parameter space. In other words: after this point is passed, the system can't return to a stable steady state (see **Figure 24**). Calculating the eigenvalues of the Jacobian at each step ensures that the iteration is terminated when one eigenvalue exceeds $\lambda = -10^{-5}$. The 2nd criterion is that all variables are positive.

3.8 References

(1) Gleizer, S.; Ben-Nissan, R.; Bar-On, Y. M.; Antonovsky, N.; Noor, E.; Zohar, Y.; Jona, G.; Krieger, E.; Shamshoum, M.; Bar-Even, A.; Milo, R. Conversion of Escherichia Coli to Generate All Biomass Carbon from CO2. *Cell* **2019**, *179* (6), 1255-1263.e12. https://doi.org/10.1016/j.cell.2019.11.009.

(2) Yim, H.; Haselbeck, R.; Niu, W.; Pujol-Baxley, C.; Burgard, A.; Boldt, J.; Khandurina, J.; Trawick, J. D.; Osterhout, R. E.; Stephen, R.; Estadilla, J.; Teisan, S.; Schreyer, H. B.; Andrae, S.; Yang, T. H.; Lee, S. Y.; Burk, M. J.; Van Dien, S. Metabolic Engineering of Escherichia Coli for Direct Production of 1,4-Butanediol. *Nat Chem Biol* **2011**, *7* (7), 445–452. https://doi.org/10.1038/nchembio.580.

(3) Wu, G.; Yan, Q.; Jones, J. A.; Tang, Y. J.; Fong, S. S.; Koffas, M. A. G. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol* **2016**, *34* (8), 652–664. https://doi.org/10.1016/j.tibtech.2016.02.010.

(4) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat Biotechnol* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.

(5) Burgard, A. P.; Pharkya, P.; Maranas, C. D. Optknock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnol Bioeng* **2003**, *84* (6), 647–657. https://doi.org/10.1002/bit.10803.

(6) Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F. J.; Doyle, J. Robustness of Cellular Functions. *Cell* **2004**, *118* (6), 675–685. https://doi.org/10.1016/j.cell.2004.09.008.

(7) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metabolic Engineering* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.

(8) He, F.; Fromion, V.; Westerhoff, H. V. (Im)Perfect Robustness and Adaptation of Metabolic Networks Subject to Metabolic and Gene-Expression Regulation: Marrying Control Engineering with Metabolic Control Analysis. *BMC Systems Biology* **2013**, *7* (1), 131. https://doi.org/10.1186/1752-0509-7-131.

(9) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metab Eng* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.

(10) Grimbs, S.; Selbig, J.; Bulik, S.; Holzhütter, H.-G.; Steuer, R. The Stability and Robustness of Metabolic States: Identifying Stabilizing Sites in Metabolic Networks. *Mol Syst Biol* **2007**, *3*, 146. https://doi.org/10.1038/msb4100186.

(11) Bogorad, I. W.; Lin, T.-S.; Liao, J. C. Synthetic Non-Oxidative Glycolysis Enables Complete Carbon Conservation. *Nature* **2013**, *502* (7473), 693–697. https://doi.org/10.1038/nature12575.

(12) Fung, E.; Wong, W. W.; Suen, J. K.; Bulter, T.; Lee, S.; Liao, J. C. A Synthetic Gene-Metabolic Oscillator. *Nature* **2005**, *435* (7038), 118–122. https://doi.org/10.1038/nature03508.

(13) Langan, R. A.; Boyken, S. E.; Ng, A. H.; Samson, J. A.; Dods, G.; Westbrook, A. M.; Nguyen, T. H.; Lajoie, M. J.; Chen, Z.; Berger, S.; Mulligan, V. K.; Dueber, J. E.; Novak, W. R. P.; El-Samad, H.; Baker, D. De Novo Design of Bioactive Protein Switches. *Nature* **2019**, *572* (7768), 205–210. https://doi.org/10.1038/s41586-019-1432-8.

(14) Farmer, W. R.; Liao, J. C. Improving Lycopene Production in Escherichia Coli by Engineering Metabolic Control. *Nat Biotechnol* **2000**, *18* (5), 533–537. https://doi.org/10.1038/75398.
(15) Zhang, F.; Carothers, J. M.; Keasling, J. D. Design of a Dynamic Sensor-Regulator System for Production of Chemicals and Fuels Derived from Fatty Acids. *Nat Biotechnol* **2012**, *30* (4), 354–359. https://doi.org/10.1038/nbt.2149.
(16) Dahl, R. H.; Zhang, F.; Alonso-Gutierrez, J.; Baidoo, E.; Batth, T. S.; Redding-Johanson, A. M.; Petzold, C. J.; Mukhopadhyay, A.; Lee, T. S.; Adams, P. D.; Keasling, J. D. Engineering Dynamic Pathway Regulation Using Stress-Response Promoters. *Nat Biotechnol* **2013**, *31* (11), 1039–1046. https://doi.org/10.1038/nbt.2689.
(17) Du, B.; Zielinski, D. C.; Kavvas, E. S.; Dräger, A.; Tan, J.; Zhang, Z.; Ruggiero, K. E.; Arzumanyan, G. A.; Palsson, B. O. Evaluation of Rate Law Approximations in Bottom-up Kinetic Models of Metabolism. *BMC Systems Biology* **2016**, *10* (1), 40. https://doi.org/10.1186/s12918-016-0283-2.
(18) Khodayari, A.; Maranas, C. D. A Genome-Scale Escherichia Coli Kinetic Metabolic Model k-Ecoli457 Satisfying Flux Data for Multiple Mutant Strains. *Nat Commun* **2016**, *7* (1), 13806. https://doi.org/10.1038/ncomms13806.
(19) Khodayari, A.; Zomorrodi, A. R.; Liao, J. C.; Maranas, C. D. A Kinetic Model of Escherichia Coli Core Metabolism Satisfying Multiple Sets of Mutant Flux Data. *Metabolic Engineering* **2014**, *25*, 50–62. https://doi.org/10.1016/j.ymben.2014.05.014.
(20) Greene, J.; Daniell, J.; Köpke, M.; Broadbelt, L.; Tyo, K. E. J. Kinetic Ensemble Model of Gas Fermenting Clostridium Autoethanogenum for Improved Ethanol Production. *Biochemical Engineering Journal* **2019**, *148*, 46–56. https://doi.org/10.1016/j.bej.2019.04.021.
(21) Sudarsan, S.; Blank, L. M.; Dietrich, A.; Vielhauer, O.; Takors, R.; Schmid, A.; Reuss, M. Dynamics of Benzoate Metabolism in Pseudomonas Putida KT2440. *Metab Eng Commun* **2016**, *3*, 97–110. https://doi.org/10.1016/j.meteno.2016.03.005.
(22) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli. *Nat Chem Biol* **2009**, *5* (8), 593–599. https://doi.org/10.1038/nchembio.186.
(23) Gerosa, L.; Haverkorn van Rijsewijk, B. R. B.; Christodoulou, D.; Kochanowski, K.; Schmidt, T. S. B.; Noor, E.; Sauer, U. Pseudo-Transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Syst* **2015**, *1* (4), 270–282. https://doi.org/10.1016/j.cels.2015.09.008.
(24) Steuer, R.; Gross, T.; Selbig, J.; Blasius, B. Structural Kinetic Modeling of Metabolic Networks. *Proc Natl Acad Sci U S A* **2006**, *103* (32), 11868–11873. https://doi.org/10.1073/pnas.0600013103.
(25) Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Res* **2002**, *30* (1), 47–49. https://doi.org/10.1093/nar/30.1.47.
(26) Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers—the Database of Key Numbers in Molecular and Cell Biology. *Nucleic Acids Res* **2010**, *38* (Database issue), D750–D753. https://doi.org/10.1093/nar/gkp889.

(27) Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **2010**, *330* (6007), 1099–1102. https://doi.org/10.1126/science.1192588.

Chapter 4 A mathematical model reveals the function of an ornithine-based allosteric activation of CarAB

Donati, S.[#]; Kuntz, M.[#]; Pahl, V.; Farke, N.; Beuter, D.; Glatter, T.; Gomes-Filho, J. V.; Randau, L.; Wang, C.-Y.; Link, H. Multi-Omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms That Buffer Decreases of Enzymes in E. Coli Metabolism. *Cell Systems* **2021**, *12* (1), 56-67.e6. https://doi.org/10.1016/j.cels.2020.10.011.

[#] Authors contributed equally

Parts of this chapter are published in Cell Systems 2021, 12 (1), 56-67.e6. My contribution to this work was the development and the analysis of the mechanistic mathematical model.

4.1 Abstract

High enzyme levels are costly, but they enable robust metabolism. Conversely, low enzyme levels are more efficient but limit metabolic flux. To understand the consequences of reducing enzyme amounts below critical concentrations, we used CRISPR interference (CRISPRi) to knockdown gene expression of CarAB, an enzyme located upstream of the branch point between arginine biosynthesis and pyrimidine biosynthesis. Following the knockdown of CarAB, we observed a 512-fold accumulation of ornithine, which is an allosteric activator of CarAB. To understand the functional implications of ornithine accumulation, we developed a mechanistic mathematical model of the branch point between arginine biosynthesis and pyrimidine biosynthesis. Simulating the knock-down of CarAB for an allosterically regulated model, and a model without allosteric regulation revealed that ornithine buffers the perturbation of CarAB and confers robustness by stabilizing downstream metabolite levels in arginine and pyrimidine biosynthesis.

4.2 Introduction

Expression levels of enzymes influence cellular metabolism and fitness. Growth of *E. coli*, for example, is affected by the abundance of single enzymes[1,2], as well as by the total mass of catabolic enzymes[3]. While studies using genetic knockouts in yeast and *E. coli*[4,5] have shown that the metabolome and transcriptome is perturbed locally[6–8], knockouts are extreme cases that are not applicable for metabolic enzymes that are essential on carbon sources such as glucose. Moreover, knockouts are static and, thus, they may reflect metabolic states that have already adapted at the level of gene expression[9]. This makes it difficult to study the immediate dynamics leading to the adapted state. Gene expression noise or mutations of genes that encode enzymes can lead to small changes in enzyme levels. However, it is still unclear how metabolism adapts to small changes of single enzymes. Control theory, for example, suggests that small changes in enzyme levels only have small and local effects on metabolism and that these changes do not propagate globally[10,11]. While this robustness of metabolism is not surprising, mostly theoretical studies examined the mechanisms that enable metabolic robustness[12,13].

Earlier studies that measured robustness against changes of enzyme abundance focused on specific pathways[14,15]. Recent approaches enabled robustness analyses by simulating strong genetic perturbations in large mechanistic models[16,17].

Strong perturbations resemble knockdowns by CRISPR interference (CRISPRi)[18], which represses transcription of a target gene with a complex of deactivated Cas9 (dCas9) and a single guide RNA (sgRNA). Since CRISPRi is inducible, it allows for time-resolved[19] studies and functional analyses of genes that are essential and, therefore, not viable in knockout libraries[20]. Thus, together with multi-omics data and mathematical models, CRISPRi based knockdowns have the potential to explain how metabolism compensates changes in enzyme levels. Here, we used CRISPRi to knockdown CarAB, an enzyme at the branch point between arginine and pyrimidine biosynthesis. We measured the metabolome and observed that ornithine, a known allosteric activator of CarAB, accumulated 512-fold. To understand if ornithine accumulation is a direct cause of the CarAB knockdown, and its functional role at the branch point, we developed a mechanistic mathematical model of the branch point. A Robustness analysis and time-course simulations of the CarAB knockdown revealed that ornithine accumulates in response to a downstream bottleneck. Allosteric activation of CarAB by ornithine alleviates the bottleneck, stabilizing pathway flux and causing end-product homeostasis.

4.3 CRISPRi knockdown of CarAB causes ornithine accumulation

To test how metabolism compensates genetic knockdowns, we used CRISPRi to perturb the carbamoyl-phosphate synthetase (CarAB) at the branch point between arginine biosynthesis and pyrimidine biosynthesis. We measured 119 intracellular metabolites by liquid chromatography tandem mass spectrometry (LC-MS/MS). Surprisingly, the substrate L-glutamine was not among the strongest changing metabolites. Instead, ornithine showed the strongest accumulation with a fold change > 500 compared to an uninduced control strain (**Figure 26a**). Ornithine is an allosteric activator of CarAB and ornithine levels in *E. coli* are 37-fold lower than the activation constant of CarAB[21]. Assuming a hyperbolic relationship between ornithine levels and CarAB activity, the initial activity of CarAB is at 3 % of its theoretical maximum activity. A 512-fold increase of ornithine should then cause a 91 % increase of activity of CarAB *in vivo*. Thus, allosteric

activation of CarAB by ornthine has the potential to buffer the CarAB knockdown (**Figure 26b**).

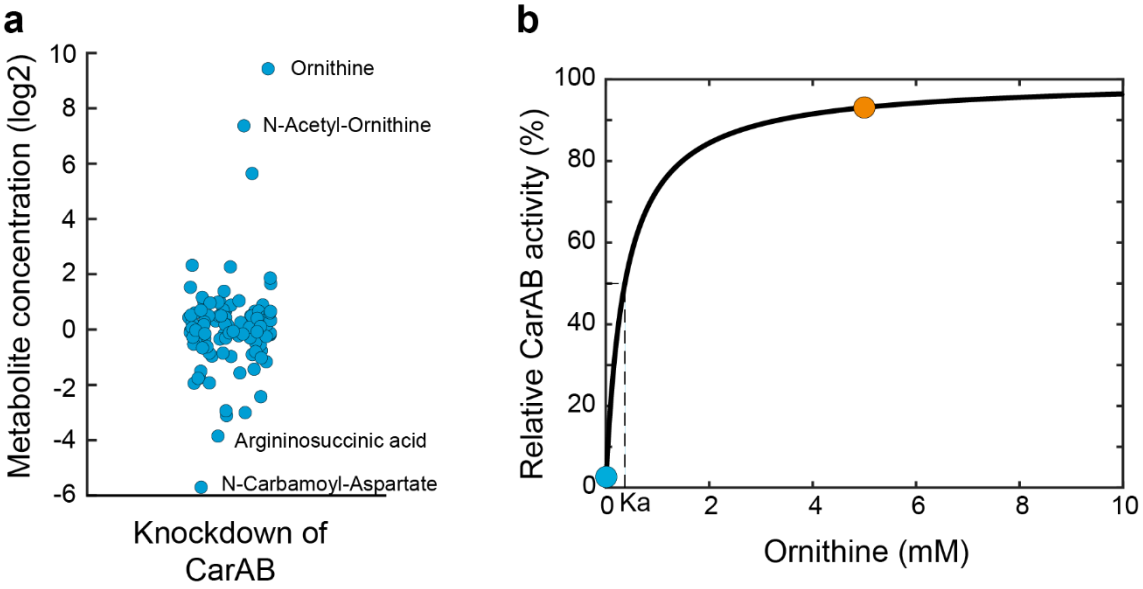


Figure 26. Knockdown of CarAB causes ornithine accumulation a) *Intracellular concentration of 119 metabolites in the 30 CRISPRi strains. Metabolite levels are shown as log2 fold-change between induced and uninduced cultures. Samples were collected after 4.5 h cultivation in 12-well plates. Data are represented as mean (n=2).* ***b)*** *Allosteric activation of CarAB by ornithine was modelled with a hyperbolic function and an activation constant (Ka) of 0.37 mM[22]. The wild-type concentration of ornithine is 0.01 mM[21] and shown as blue dot (3% activity). The concentration of ornithine increased 512-fold in the CarAB knockdown (orange dot, 94% activity). Parts of this figure are published in Cell Systems 2021, 12 (1), 56-67.e6.*

In conclusion, we performed a knockdown of CarAB and measured metabolite changes of 119 metabolites with LC-MS/MS. Ornithine accumulated 512-fold and literature ornithine concentrations and activation constants of CarAB predicted that this accumulation should increase CarAB activity by 91 %. To provide further mechanistic insights, we next developed a small mechanistic model of the branch between arginine biosynthesis and pyrimidine biosynthesis.

4.4 Feedback activation of CarAB by ornithine stabilizes pathway end-products in arginine and pyrimidine biosynthesis

To study the role of the allosteric activation of CarAB by ornithine, we developed a small mechanistic model of the arginine and pyrimidine branch point. The model consists of six metabolic reactions (r1-r6) and four metabolites (orn: ornithine, cbp: carbamoyl

phosphate, arg: arginine, utp/ctp: uridine-triphosphate / cytidine-triphosphate). Reaction 2 includes an additional transcriptional layer in which we simulate production of CarAB and its dilution by growth. Moreover, reaction 2 is allosterically activated by ornithine, and we model the activation with a power-law term. The metabolic end products arginine and UTP/CTP are then consumed for growth and we assumed that both end products contribute equally to the growth rate. The model stoichiometry is shown in **Figure 27a**. We then set up an allosteric model (with ornithine feedback activation) and a dysregulated model (without ornithine feedback activation). Kinetic parameters of the model were randomly sampled 1,000 times from physiologically meaningful ranges based on *in vitro* parameters. Note that both models reach the same stable steady state flux and share the same kinetic parameters (except for the power-law term).

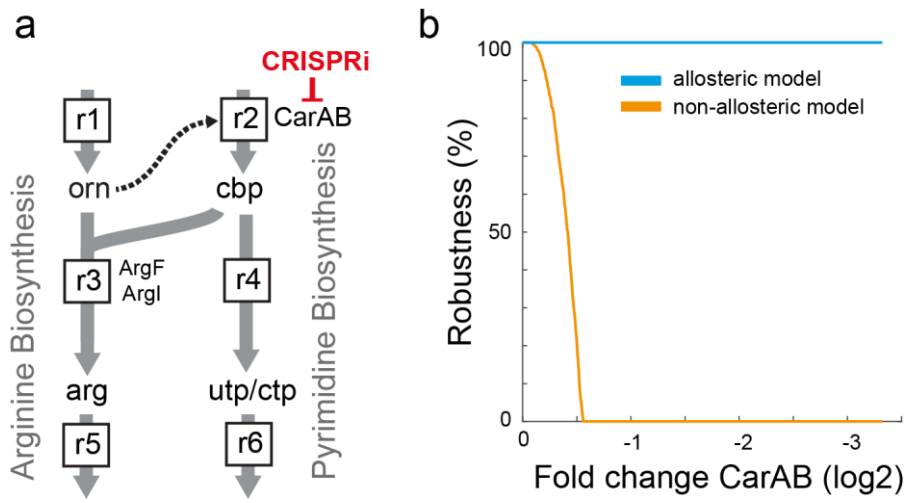


Figure 27. Model structure and robustness analysis a) *Stoichiometry of the kinetic model of the branch point between arginine and pyrimidine biosynthesis. Reactions (r1-r6) are solid arrows. The allosteric ornithine activation of r2 is a dashed arrow. The perturbation of CarAB by CRISPRi is a red blunt arrow.* ***b)*** *Robustness analysis of an allosteric(blue) and a non-allosteric model(orange). Robustness is defined as the fraction of stable parameter sets at each perturbation step (1,000 parameter sets in total). Parts of this figure are published in Cell Systems 2021, 12 (1), 56-67.e6.*

Analysis of the literature parameters indicated that the allosteric activation of CarAB by ornithine should increase CarAB activity. An initial robustness analysis of the allosteric model and the non-allosteric model indicated that allosteric activation of CarAB is crucial for robustness of the branch point. To perform the robustness analysis, we reduced the production rate of CarAB in both models until instabilities occured. Repeating this

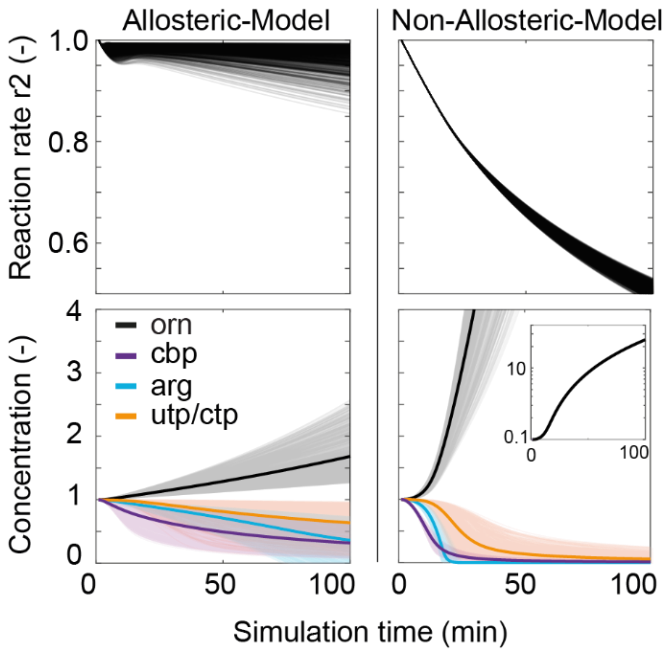


Figure 28. Time course simulation of the CarAB knockdown. *Simulation results of the allosteric model and the non-allosteric model with 1,000 parameter sets (thin lines). Thick lines are the average of 1,000 simulations. Shown are the simulated reaction rate of r2 and metabolite dynamics of ornithine (orn, black), carbamoyl phosphate (cbp, purple), arginine (arg, blue) and UTP/CTP (orange). CRISPRi was simulated by setting the expression rate of CarAB to zero at t = 0 min. The insert shows the full range average ornithine levels in the non-allosteric model. This figure is published in Cell Systems 2021, 12 (1), 56-67.e6*

process for all 1,000 parameter sets, we defined robustness as the fraction of stable parameter sets at each perturbation step. While the allosteric model maintained maximum robustness up to a 10-fold downregulation of CarAB, the non-allosteric model could not even tolerate a two-fold downregulation, indicating a severe lack of robustness (**Figure 27**). This analysis showed that the allosteric model is far more robust against a bottleneck in CarAB than the non-allosteric model.

To better understand cause and effect of the CRISPRi knockdown of *CarAB*, we next simulated the time course of the knockdown for both models and all 1,000 parameter sets. Again, the allosteric model was more robust against the CarAB knockdown than the non-allosteric model (**Figure 28**). Especially fluxes remained relatively constant in the allosteric model: 796 of the 1,000 simulations maintained 95% of the initial steady-state flux. In contrast, the flux in the non-allosteric model decreased continuously to about 50% of the initial steady state. Moreover, concentrations of the end products, arginine and UTP/CTP, were more stable in the allosteric model than in the non-allosteric model. Out of all metabolites, the levels of carbamoyl phosphate drop first,

followed by the accumulation of ornithine. This effect is stronger in the non-allosteric model, indicating that ornithine accumulates because of a bottleneck in reaction 3, and not as a direct response to buffer the knockdown of CarAB.

The simulations indicate that ornithine does not accumulate because of the perturbation of CarAB, but due to the bottleneck in reaction 3. The allosteric activation of CarAB by ornithine then alleviates the bottleneck, thus minimizing perturbations to metabolic flux and end products in arginine and pyrimidine nucleotide biosynthesis.

To validate the model predictions, we measured metabolites in the CarAB knockdown dynamically with LC-MS/MS (**Figure 29**). Consistent with the simulation results, ornithine already increased 40 min after induction of the CarAB knockdown, while the end products, arginine, CTP, and UTP remained relatively constant for at least 2 h (Carbamoyl phosphate is volatile and could not be measured): The fast response of ornithine shows that the CarAB knockdown perturbs the arginine-pyrimidine branch point early after induction of CRISPRi. However, the perturbation did not propagate into the end products of both pathways. Thus, the combination of a mechanistic model and dynamic metabolite data provided additional evidence that ornithine buffers the CarAB knockdown.

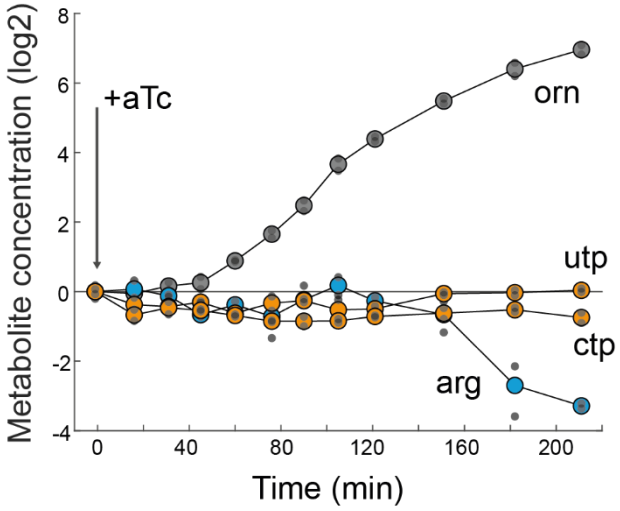


***Figure 29. Time course measurements of metabolite concentrations**. Measured concentration of orn, arg, UTP, and CTP in the CarAB knockdown. Metabolite levels are normalized to the time point before induction. The culture was induced with aTc at t = 0 min. Small gray dots are measurements in n = 2 cultures and large colored dots are the mean. This figure is published in Cell Systems 2021, 12 (1), 56-67.e6*

4.5 Discussion

Robustness is an emergent property of metabolism and allows systems to evolve and maintain functionality in face of perturbations[23,24]. To achieve robustness, cells use feedback control to dynamically adapt their metabolism to changes in metabolite and enzyme levels[23,24]. However, due to lack of time-resolved multi-omics data and mild genetic perturbations, there is only little information on how exactly feedback mechanisms like allosteric regulation achieve robustness on a system level. To study how cells adapt to perturbations metabolically, mathematical models have been used extensively[25–27]. While metabolic control analysis has predicted that metabolism is insensitive to perturbations of enzyme levels[10–12], analyses of mechanistic models have shown that metabolic networks, depending on the network motif, don't necessarily tolerate such perturbations leading to system failure[16,17].

In this chapter, we used CRISPRi to knockdown the enzyme CarAB at the branch-point between arginine biosynthesis and pyrimidine biosynthesis. Analysis of metabolite data revealed that the CarAB knockdown caused 512-fold accumulation of ornithine, a known allosteric activator of CarAB. To simulate the knockdown of CarAB and to understand the compensatory response of the pathway, we developed a mechanistic model of the arginine-pyrimidine branch point. A robustness analysis and time-course simulations for an allosteric and non-allosteric model indicated that the knockdown leads to a downstream bottleneck that causes ornithine accumulation. Subsequent ornithine-activation of CarAB alleviates the bottleneck and stabilizes biosynthetic pathway end-products.

Here, we have demonstrated that allosteric activation plays a crucial role in buffering decreases of enzyme levels. It seems likely that structurally similar pathways share this regulatory motif to solve downstream bottlenecks. Moreover, this case study indicates that functional metabolite-protein interactions could be identified by systematically perturbating enzyme levels, since perturbations can cause strong accumulation or depletion of regulatory metabolites[28].

4.6 Methods

Structure of the mechanistic model and time course simulation

The stoichiometry of the model is shown in **Figure 27a**. Mass balancing yields a system of ordinary differential equations (ODEs), F, that is a temporal function of the state variables x and the kinetic parameters p:

$$F(x,p) = \frac{dx}{dt} = \begin{cases} \frac{dorn}{dt} = r_1 - r_3 \\ \frac{dcbp}{dt} = r_2 - r_3 - r_4 \\ \frac{darg}{dt} = r_3 - \alpha_1\,\mu \\ \frac{dutp}{dt} = r_4 - \alpha_2\,\mu \\ \frac{de_2}{dt} = -\mu \cdot e_2 \end{cases} \quad (4.1)$$

The six reactions (r1, r2, r3, r4, μ1, μ2) are described by the following kinetic equations:

The influx into the arginine pathway r1 is constant:

$$r1 = k_{cat1} \cdot e_1 \quad (4.2)$$

Allosteric activation of reaction r2 by ornithine follows a power-law function:

$$r2 = k_{cat2} \cdot e_2 \cdot \left(\frac{orn}{orn_{SS}}\right)^{K_2} \quad (4.3)$$

where orn_{SS} is the steady state ornithine concentration. Reaction r3 corresponds to the ornithine carbamoyl transferase, which was shown to follow an ordered bi-bi enzyme mechanism in which initial carbamoyl phosphate binding induces a conformational change that allows for binding of ornithine[29]. We assumed that the phosphate can be neglected and change the mechanism to an ordered Bi-Uni mechanism. Ordered and non-ordered bi-uni mechanisms cannot be distinguished in the absence of product inhibition[30]. Thus, reaction r3 follows a non-ordered bi-uni mechanism:

$$r3 = k_{cat3} \cdot e_3 \; \frac{1}{\left(1 + \frac{K_{m,orn} \cdot K_{m,cbp}}{orn \cdot cbp} + \frac{K_{m,orn}}{orn} + \frac{K_{m,cbp}}{cbp}\right)} \quad (4.4)$$

Reaction r4 follows Michaelis-Menten kinetics:

$$r4 = k_{cat4} \cdot e_4 \cdot \frac{cbp}{cbp + K_4} \quad (4.5)$$

The growth rate μ depends on $\mu 1$ and $\mu 2$, which follow Michaelis-Menten kinetics:

$$\mu_1 = \mu_{1,max} \cdot \frac{arg}{arg + K_{\mu 1}} \quad (4.6)$$

$$\mu_2 = \mu_{2,max} \cdot \frac{utp}{utp + K_{\mu 2}} \quad (4.7)$$

$$\mu = mean(\mu 1, \mu 2) \quad (4.8)$$

In total, the model includes 14 kinetic parameters k_{cat1}, k_{cat2}, k_{cat3}, k_{cat4}, K_2, $K_{m,orn}$, $K_{m,cbp}$, K_4, $K_{\mu 1}$, $K_{\mu 2}$, μ_{max1}, μ_{max2}, α_1 and α_2. The ensemble modelling approach[31] was used to account for uncertainties in kinetic parameters.

First, a steady state flux distribution was calculated by FBA that is common for all subsequent parameter sets (r1 = 0.958 mM min^{-1}, r2 = 1.425 mM min^{-1}, r3 = 0.958 mM min^{-1}, r4 = 0.467 mM min^{-1}, $\mu 1$ = 0.958 mM min^{-1}, $\mu 2$ = 0.467 mM min^{-1}). Arginine and UTP efflux ($\mu 1$ and $\mu 2$) were calculated as the product of their biomass coefficients (α1 = 95.8 mM, α2 = 46.7 mM) and the growth rate on glucose (μ = 0.01 min^{1}).

Binding constants (K-values) and metabolite concentrations (Ornithine = 0.01 mM, UMP = 0.50 mM, Arginine = 0.138 mM[21]) were obtained from literature and Cbp concentration was set to 1 mM. The enzyme concentrations were set to 1 mM. The binding constants were sampled 1,000 times from 10-fold intervals based on literature values ($K_{m,orn}$ = 0.32 mM (argF/I, Ecocyc[32]), Km,cbp = 0.36 mM (argF/I, Ecocyc), K4 = 0.028 mM (Brenda[33] ID: 696699), $K_{\mu 2}$ = 0.05 mM (pyrH, Ecocyc)). The power-law term K_2 was sampled between 1 and 4 in the regulated model[25] and was set to zero in the dysregulated model. $K_{\mu 1}$ was fixed to 10^{-5} mM.

With the ensemble modelling approach, the system is initially set into a steady state. To test stability of the steady states, eigenvalues of the Jacobian matrix were calculated, and tested if all eigenvalues are negative ($\lambda < -10^{-5}$). The procedure was repeated until 1,000 stable steady states were achieved. The perturbation by CRISPRi was then simulated for all stable models by setting the expression rate of e2 to zero:

$$\frac{de_2}{dt} = -e_2 \, \mu \quad (4.9)$$

Robustness Analysis

We first removed enzyme production and dilution from the model (**Equation 4.9**). To determine robustness, we used a numerical parameter continuation method to iteratively decreased e2 in both models until instabilities occured. The method is based on finding a connected path of steady state concentrations (x_{ss}: steady state concentration vector containing: orn, cbp, arg, utp/ctp) as a parameter p is varied. As the system is in steady state it follows that:

$$\frac{dx}{dt} = F(x_{ss}, p) = 0 \tag{4.10}$$

The derivative of $F(x_{ss},p)$ with respect to the parameters is also zero:

$$\frac{dF(x_{ss}, p)}{dp} = \frac{\delta F}{\delta x_{ss}} \cdot \frac{dx_{ss}}{dp} + \frac{\delta F}{\delta p} = 0 \tag{4.11}$$

After rearranging **Equation 4.11**, **Equation 4.12** is obtained:

$$\frac{dx_{ss}}{dp} = -\left(\frac{\delta F}{\delta x_{ss}}\right)^{-1} \cdot \frac{\delta F}{\delta p} \tag{4.12}$$

which describes the changes in the steady-state concentrations as a kinetic parameter is varied iteratively. The iteration stops when one of the following two stability criteria is no longer fulfilled. 1st criterion: all real parts of the eigenvalues of the system's Jacobian need to be negative. In **Equation 4.12** the inverse of the Jacobian Matrix ($\delta F/\delta x_{SS}$) is required. The inversion is only possible if the matrix is regular. Once an eigenvalue reaches zero, the Jacobian becomes singular and matrix inversion is no longer possible. This bifurcation point defines the boundary between the stable and unstable parameter space. In other words: after this point is passed, the system can't return to a stable steady state. Calculating the eigenvalues of the Jacobian at each step ensures that the iteration is terminated when one eigenvalue exceeds $\lambda = -10^{-5}$. The 2nd criterion is that all variables are positive.

4.7 References

(1) Li, G.-W.; Burkhardt, D.; Gross, C.; Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **2014**, *157* (3), 624–635. https://doi.org/10.1016/j.cell.2014.02.033.

(2) Dekel, E.; Alon, U. Optimality and Evolutionary Tuning of the Expression Level of a Protein. *Nature* **2005**, *436* (7050), 588–592. https://doi.org/10.1038/nature03842.
(3) You, C.; Okano, H.; Hui, S.; Zhang, Z.; Kim, M.; Gunderson, C. W.; Wang, Y.-P.; Lenz, P.; Yan, D.; Hwa, T. Coordination of Bacterial Proteome with Metabolism by Cyclic AMP Signalling. *Nature* **2013**, *500* (7462), 301–306. https://doi.org/10.1038/nature12446.
(4) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Mol Syst Biol* **2006**, *2*, 2006.0008. https://doi.org/10.1038/msb4100050.
(5) Giaever, G.; Chu, A. M.; Ni, L.; Connelly, C.; Riles, L.; Véronneau, S.; Dow, S.; Lucau-Danila, A.; Anderson, K.; André, B.; Arkin, A. P.; Astromoff, A.; El-Bakkoury, M.; Bangham, R.; Benito, R.; Brachat, S.; Campanaro, S.; Curtiss, M.; Davis, K.; Deutschbauer, A.; Entian, K.-D.; Flaherty, P.; Foury, F.; Garfinkel, D. J.; Gerstein, M.; Gotte, D.; Güldener, U.; Hegemann, J. H.; Hempel, S.; Herman, Z.; Jaramillo, D. F.; Kelly, D. E.; Kelly, S. L.; Kötter, P.; LaBonte, D.; Lamb, D. C.; Lan, N.; Liang, H.; Liao, H.; Liu, L.; Luo, C.; Lussier, M.; Mao, R.; Menard, P.; Ooi, S. L.; Revuelta, J. L.; Roberts, C. J.; Rose, M.; Ross-Macdonald, P.; Scherens, B.; Schimmack, G.; Shafer, B.; Shoemaker, D. D.; Sookhai-Mahadeo, S.; Storms, R. K.; Strathern, J. N.; Valle, G.; Voet, M.; Volckaert, G.; Wang, C.; Ward, T. R.; Wilhelmy, J.; Winzeler, E. A.; Yang, Y.; Yen, G.; Youngman, E.; Yu, K.; Bussey, H.; Boeke, J. D.; Snyder, M.; Philippsen, P.; Davis, R. W.; Johnston, M. Functional Profiling of the Saccharomyces Cerevisiae Genome. *Nature* **2002**, *418* (6896), 387–391. https://doi.org/10.1038/nature00935.
(6) Fuhrer, T.; Zampieri, M.; Sévin, D. C.; Sauer, U.; Zamboni, N. Genomewide Landscape of Gene-Metabolome Associations in Escherichia Coli. *Mol Syst Biol* **2017**, *13* (1), 907. https://doi.org/10.15252/msb.20167150.
(7) Mülleder, M.; Calvani, E.; Alam, M. T.; Wang, R. K.; Eckerstorfer, F.; Zelezniak, A.; Ralser, M. Functional Metabolomics Describes the Yeast Biosynthetic Regulome. *Cell* **2016**, *167* (2), 553-565.e12. https://doi.org/10.1016/j.cell.2016.09.007.
(8) Kemmeren, P.; Sameith, K.; van de Pasch, L. A. L.; Benschop, J. J.; Lenstra, T. L.; Margaritis, T.; O'Duibhir, E.; Apweiler, E.; van Wageningen, S.; Ko, C. W.; van Heesch, S.; Kashani, M. M.; Ampatziadis-Michailidis, G.; Brok, M. O.; Brabers, N. A. C. H.; Miles, A. J.; Bouwmeester, D.; van Hooff, S. R.; van Bakel, H.; Sluiters, E.; Bakker, L. V.; Snel, B.; Lijnzaad, P.; van Leenen, D.; Groot Koerkamp, M. J. A.; Holstege, F. C. P. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. *Cell* **2014**, *157* (3), 740–752. https://doi.org/10.1016/j.cell.2014.02.054.
(9) Hosseini, S.-R.; Wagner, A. Genomic Organization Underlying Deletional Robustness in Bacterial Metabolic Systems. *Proc Natl Acad Sci U S A* **2018**, *115* (27), 7075–7080. https://doi.org/10.1073/pnas.1717243115.
(10) Kacser, H.; Burns, J. A. The Control of Flux. *Biochem Soc Trans* **1995**, *23* (2), 341–366. https://doi.org/10.1042/bst0230341.
(11) Levine, E.; Hwa, T. Stochastic Fluctuations in Metabolic Pathways. *Proc Natl Acad Sci U S A* **2007**, *104* (22), 9224–9229. https://doi.org/10.1073/pnas.0610987104.

(12) Chandra, F. A.; Buzi, G.; Doyle, J. C. Glycolytic Oscillations and Limits on Robust Efficiency. *Science* **2011**, *333* (6039), 187–192. https://doi.org/10.1126/science.1200705.
(13) The Stability and Robustness of Metabolic States: Identifying Stabilizing Sites in Metabolic Networks. *Molecular Systems Biology* **2007**, *3* (1), 146. https://doi.org/10.1038/msb4100186.
(14) Fendt, S.-M.; Buescher, J. M.; Rudroff, F.; Picotti, P.; Zamboni, N.; Sauer, U. Tradeoff between Enzyme and Metabolite Efficiency Maintains Metabolic Homeostasis upon Perturbations in Enzyme Capacity. *Mol Syst Biol* **2010**, *6*, 356. https://doi.org/10.1038/msb.2010.11.
(15) Tanner, L. B.; Goglia, A. G.; Wei, M. H.; Sehgal, T.; Parsons, L. R.; Park, J. O.; White, E.; Toettcher, J. E.; Rabinowitz, J. D. Four Key Steps Control Glycolytic Flux in Mammalian Cells. *Cell Syst* **2018**, *7* (1), 49-62.e8. https://doi.org/10.1016/j.cels.2018.06.003.
(16) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metabolic Engineering* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.
(17) Lafontaine Rivera, J. G.; Lee, Y.; Liao, J. C. An Entropy-like Index of Bifurcational Robustness for Metabolic Systems. *Integr Biol (Camb)* **2015**, *7* (8), 895–903. https://doi.org/10.1039/c4ib00257a.
(18) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.
(19) Camsund, D.; Lawson, M. J.; Larsson, J.; Jones, D.; Zikrin, S.; Fange, D.; Elf, J. Time-Resolved Imaging-Based CRISPRi Screening. *Nat Methods* **2020**, *17* (1), 86–92. https://doi.org/10.1038/s41592-019-0629-y.
(20) Peters, J. M.; Colavin, A.; Shi, H.; Czarny, T. L.; Larson, M. H.; Wong, S.; Hawkins, J. S.; Lu, C. H. S.; Koo, B.-M; Marta, E.; Shiver, A. L.; Whitehead, E. H.; Weissman, J. S.; Brown, E. D.; Qi, L. S.; Huang, K. C.; Gross, C. A. A Comprehensive, CRISPR-Based Functional Analysis of Essential Genes in Bacteria. *Cell* **2016**, *165* (6), 1493–1506. https://doi.org/10.1016/j.cell.2016.05.003.
(21) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli. *Nat Chem Biol* **2009**, *5* (8), 593–599. https://doi.org/10.1038/nchembio.186.
(22) Bueso, J.; Cervera, J.; Fresquet, V.; Marina, A.; Lusty, C. J.; Rubio, V. Photoaffinity Labeling with the Activator IMP and Site-Directed Mutagenesis of Histidine 995 of Carbamoyl Phosphate Synthetase from Escherichia Coli Demonstrate That the Binding Site for IMP Overlaps with That for the Inhibitor UMP. *Biochemistry* **1999**, *38* (13), 3910–3917. https://doi.org/10.1021/bi982871f.
(23) Kitano, H. Biological Robustness. *Nat Rev Genet* **2004**, *5* (11), 826–837. https://doi.org/10.1038/nrg1471.
(24) Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F. J.; Doyle, J. Robustness of Cellular Functions. *Cell* **2004**, *118* (6), 675–685. https://doi.org/10.1016/j.cell.2004.09.008.

(25) Grimbs, S.; Selbig, J.; Bulik, S.; Holzhütter, H.-G.; Steuer, R. The Stability and Robustness of Metabolic States: Identifying Stabilizing Sites in Metabolic Networks. *Mol Syst Biol* **2007**, *3*, 146. https://doi.org/10.1038/msb4100186.
(26) Kremling, A.; Bettenbrock, K.; Gilles, E. D. A Feed-Forward Loop Guarantees Robust Behavior in Escherichia Coli Carbohydrate Uptake. *Bioinformatics* **2008**, *24* (5), 704–710. https://doi.org/10.1093/bioinformatics/btn010.
(27) Link, H.; Kochanowski, K.; Sauer, U. Systematic Identification of Allosteric Protein-Metabolite Interactions That Control Enzyme Activity in Vivo. *Nat Biotechnol* **2013**, *31* (4), 357–361. https://doi.org/10.1038/nbt.2489.
(28) Donati, S.; Kuntz, M.; Pahl, V.; Farke, N.; Beuter, D.; Glatter, T.; Gomes-Filho, J. V.; Randau, L.; Wang, C.-Y.; Link, H. Multi-Omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms That Buffer Decreases of Enzymes in E. Coli Metabolism. *Cell Systems* **2021**, *12* (1), 56-67.e6. https://doi.org/10.1016/j.cels.2020.10.011.
(29) Kuo, L. C.; Seaton, B. A. X-Ray Diffraction Analysis on Single Crystals of Recombinant Escherichia Coli Ornithine Transcarbamoylase. *J Biol Chem* **1989**, *264* (27), 16246–16248.
(30) Wang, Y.; Mittermaier, A. K. Characterizing Bi-Substrate Enzyme Kinetics at High Resolution by 2D-ITC. *Anal. Chem.* **2021**, *93* (37), 12723–12732. https://doi.org/10.1021/acs.analchem.1c02705.
(31) Tran, L. M.; Rizk, M. L.; Liao, J. C. Ensemble Modeling of Metabolic Networks. *Biophys J* **2008**, *95* (12), 5606–5617. https://doi.org/10.1529/biophysj.108.135442.
(32) *The EcoCyc Database - PubMed*. https://pubmed.ncbi.nlm.nih.gov/30406744/ (accessed 2022-07-10).
(33) Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Res* **2002**, *30* (1), 47–49. https://doi.org/10.1093/nar/30.1.47.

Chapter 5 A mathematical model of glycolysis predicts sustained pyruvate oscillations

Shuangyu Bi, Manika Kargeti, Remy Colin, Niklas Farke, Hannes Link, and Victor Sourjik

Parts of this chapter are part of the manuscript “Dynamic fluctuations in a bacterial metabolic network” that is currently under revision in Nature Communications. My contribution to this work was the design of the mathematical models and the analysis of the mathematical models.

5.1 Abstract

Modelling approaches based on network stoichiometry often assume that exponentially growing bacteria are in steady state. That means that metabolite concentrations are time invariant. Perturbations, however, can disturb this balance, causing metabolite accumulation or depletion. An intricate network of feedback regulations allows the cell to adapt to these changes by adjusting enzyme activity and enzyme abundance. This regulation network renders cells robust against certain perturbations. However, robustness against certain perturbations often causes fragility against other perturbations, for example leading to sustained oscillations of metabolite levels. Here, we developed a small mechanistic model of glycolysis that includes the phosphotransferase system and three of the most important allosteric regulations. The model indicated that feedback regulation and enzyme saturation are important factors that can cause sustained pyruvate oscillations with an average period of three minutes.

5.2 Introduction

Robustness of biological systems allows cells to maintain homeostasis of metabolite end-products, which are crucial for keeping stable growth rates[1–3]. A key assumption of most deterministic, stoichiometry-based modelling approaches is that input and output of a system are balanced and in steady state during exponential growth[4–7]. While metabolism is often assumed to be in homeostasis, cells frequently encounter perturbations, and it is not always clear how metabolism responds to these perturbations to achieve homeostasis. Allosteric regulation allows metabolism to sense changes in metabolite concentrations and adjust enzymatic reaction rates to direct the system back into a steady state[8]. Yet, there is little experimental information on how metabolite concentrations change dynamically between states. Although experimental methods like real-time metabolomics[9] have shed light on metabolite dynamics, it is more common to use mathematical models to study these effects.

Glycolysis is one of the most important and most studied catabolic pathways[10–13]. It is a tightly regulated autocatalytic pathway that provides metabolite precursors for many anabolic metabolic pathways. Due to its autocatalytic nature, glycolysis requires

phosphoenolpyruvate and ATP to produce a net gain of both metabolites. Phosphoenolpyruvate is required for uptake of glucose by the phosphotransferase system (PTS), and ATP is required for the phosphorylation of fructose-6-phosphate to fructose 1,6-bisphosphate. Moreover, feed forward activation of the pyruvate kinase by fructose 1,6-bisphosphate, and the feedback activation of the fructose-1,6-bisphosphatase by pyruvate were shown to be instrumental for robustness of glycolysis, and to switch between gluconeogenesis and glycolysis[9,14]. The complex and intertwined structure of glycolysis prohibits an intuitive understanding of glycolysis, which is crucial for many applications, including metabolic engineering[11].

However, while its complex regulation enables robustness and flexibility, it also renders glycolysis more fragile against unexpected perturbations. For example, in **Chapter 3** we showed that native transcriptional regulation of *E. coli* causes system failure when flux is drained from upper glycolysis to produce glycerol[11]. For other models it was even shown that glycolysis metabolites are susceptible to oscillations, and that these oscillations are a consequence of the many feedback regulations and autocatalytic cycles as part of the fundamental trade-off between robustness and efficiency in dynamic systems[13,15,16]. However, while these models predicted oscillations and system failure, the autocatalytic properties of the phosphotransferase system, as well as many allosteric regulations, were not considered.

Here, we developed a mechanistic mathematical model of glycolysis to study how pyruvate, one of the end-products of glycolysi[17], responds to perturbations of the glucose uptake rate. The model included the most important allosteric regulations of glycolysis and predicted that the interplay between allosteric inhibition, allosteric activation, and enzyme saturation of key enzymes can cause sustained pyruvate oscillations with an average period of three minutes.

5.3 A mathematical model of *E: coli* glycolysis predicts sustained pyruvate oscillations

To test the likelihood of oscillations in the levels of glycolysis metabolites, we developed a small mechanistic model of *E. coli* glycolysis (**Figure 30**). The model consists of four

metabolites and six metabolic reactions that were simulated with Michaelis-Menten and Hill kinetics. From a multitude of allosteric metabolite-enzyme interactions that regulate the activity of glycolysis enzymes, three of the most relevant ones were included in our model. The first interaction is the feedforward activation of pyruvate kinase (PYK) by fructose-1,6-biphosphate (FBP), which plays an important role for glycolysis flux regulation in *E. coli*[18]. The other two interactions represent negative feedbacks from the phosphoenolpyruvate (PEP) levels to the interconversion of hexose-phosphate and FBP, by respectively inhibiting phosphofructokinase (PFK) and activating fructose-1,6-bisphosphatase (FBPase), which together regulate the PFK-FBPase cycle[19].

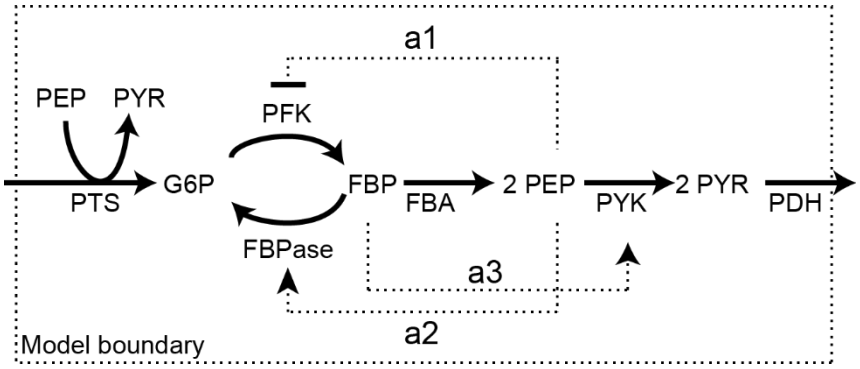


***Figure 30**. **Structure and stoichiometry of a mechanistic model of glycolysis**. The outer dotted line is the model boundary. Solid arrows are reactions and dotted arrows are allosteric interactions of metabolites with enzymes. G6P: glucose-6-phosphate, FBP: fructose-1,6-bisphosphate, PEP: phosphoenolpyruvate, PYR: pyruvate, PFK: phosphofructokinase, FBPase: fructose-1,6-bisphosphatase, PYK: pyruvate kinase, FBA: fructose-bisphosphate aldolase, PTS: phosphotransferase system, PDH: pyruvate dehydrogenase.*

As a starting point for analysis of the model, we fixed the glycolytic flux to a constant value and randomly sampled all model parameters (maximal rates and binding constants) 2,000 times from a log-uniform distribution such that the model was at steady state. In total, 50 % of the 2,000 parameter sets achieved a stable steady state, indicating that the structure alone does not guarantee stability and that parameter values of the correct magnitude are crucial. Next, we perturbed this steady state by decreasing the glucose uptake rate by 5 % and analysed changes in pyruvate levels (downshift, **Figure 31a**). A forward Fourier transformation of the time dependent pyruvate levels revealed sustained oscillations of pyruvate concentrations for 393 of the tested 2,000 parameter sets (20%).

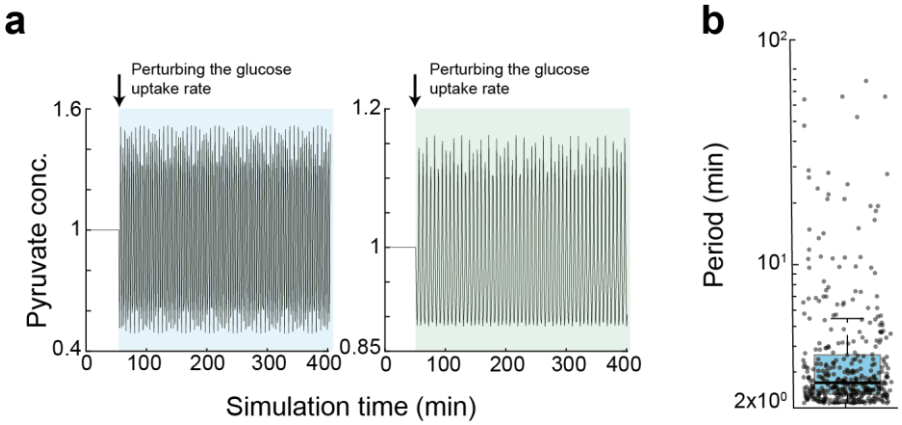


***Figure 31. Simulated Pyruvate oscillations**. **a)** Two examples of simulated pyruvate concentrations with two different parameter sets that produce pyruvate oscillations. The model was initially at steady state and at t = 50 min the glucose uptake rate (Vmax of PTS) was decreased by 5 %. **b)** Boxplot showing the distribution of the periods of 393 simulations with oscillating pyruvate levels. Each black dot corresponds to a different parameter set.*

The typical period of oscillations across these 393 simulations was several minutes (three minutes on average), although some parameter sets caused faster or slower oscillations (**Figure 31b**). However, none of the parameter sets that led to oscillations had a stable steady state. To identify parameters that favour pyruvate oscillations, we compared the distribution of individual model parameters across all sets with those sets that yielded oscillations (**Figure 32**). This showed that oscillations are generally favoured by high affinity of pyruvate dehydrogenase (PDH) for pyruvate (low Km5) and by the positive allosteric regulation of PYK by FBP (a3). Lower cooperativity of PFK (n1) and negative regulation of its activity by PEP (a1) also promoted oscillations.

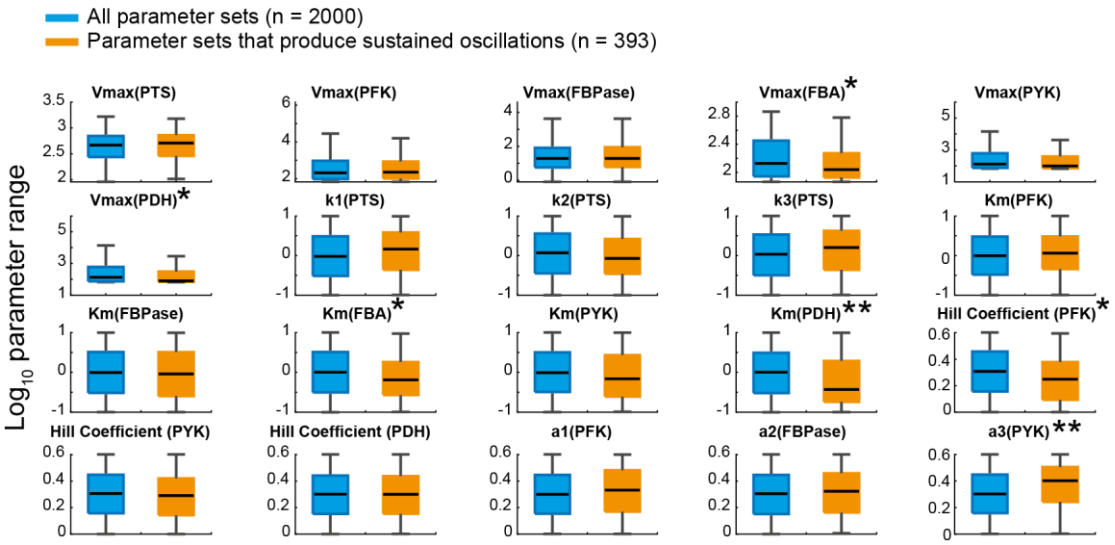


Figure 32. Analysis of model parameter sets for their propensity to produce pyruvate *oscillations. Comparison of parameter values in 393 parameter sets that led to oscillations with all 2,000 parameter sets that were randomly sampled from a log-uniform distribution, with numbering of individual reactions as in the panel. Asterisks denote parameters that affect the model's propensity to produce pyruvate oscillations (10^{-10} < p-value < 10^{-5}, a = 0.01). Double asterisks denote parameters that strongly affect the model's propensity to produce pyruvate oscillations (p-value < 10^{-10}, a = 0.01).*

Finally, we wondered how the system would respond to a 10-fold upshift of a low glucose uptake rate that resembles the shift from glucose starvation to growth. In contrast to the 5 % downshift, pyruvate levels first showed a strong increase before stabilizing and oscillating around the new steady state (**Figure 33b**). In contrast to the small perturbation of the glucose uptake rate, a 10-fold upshift caused only oscillations in 226 out of 10,000 parameter sets (2 %), while the average period of the oscillations was comparable between both simulations (**Figure 33c**). Consistent with the 5 % downshift, the allosteric feedforward activation of PYK (a3), the cooperativity of PFK (n1) and the substrate affinity of PDH for pyruvate (Km5) were also causing oscillations in the 10-fold upshift model (**Figure 33d**). Thus, the stoichiometry and the kinetics of glycolysis can produce oscillations of intracellular metabolites on the time scale of several minutes (three minutes on average) in a broad range of parameter values and for different initial perturbations (upshift and downshift of glycolytic flux).

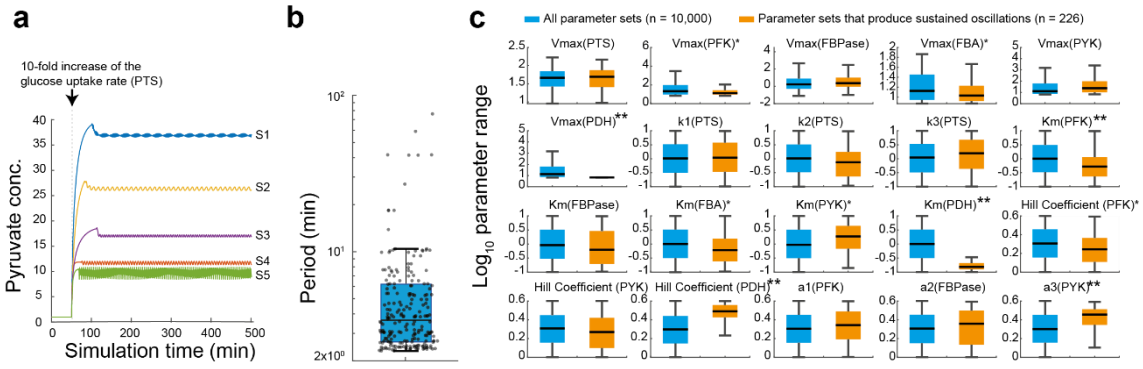


Figure 33*. *Response of the metabolic model to a 10-fold upshift in glucose uptake rate*. *a) *Five examples of simulated pyruvate concentrations with five different parameter sets (S1-S5) that produce oscillations. The model was initially at steady state and at t = 50 min with a flux of 6.66 mmol l-1min-1 when the glucose uptake rate (Vmax of PTS) was increased 10-fold.* ***b)*** *Boxplot showing the distribution of the periods of 226 simulations with oscillating pyruvate levels. Each black dot corresponds to a different parameter set.* ***c)*** *Comparison of parameter values in 226 sets that led to oscillations with all 10,000 parameter sets that were random sampled, with numbering of individual reactions as in panel. Asterisks denote parameters that affect the model's propensity to produce pyruvate oscillations (10^{-10} < p-value < 10^{-5}, a = 0.01). Double asterisks denote parameters that strongly affect the model's propensity to produce pyruvate oscillations (p-value < 10^{-10}, a = 0.01).*

5.4 Discussion

Glycolysis is one of the most important catabolic pathways and was subject to many experimental and theoretical studies[9–12,19,20]. Yet, recent studies have shown that glycolysis is far from solved, and that further knowledge will greatly aid metabolic engineering efforts[11]. Here, we developed a small mechanistic model of glycolysis, which showed that allosteric regulation and enzyme saturation can cause sustained oscillations of pyruvate levels. This result agrees with previous studies that highlighted the propensity of glycolysis to produce oscillations[13,15,16]. In contrast to those model analyses, we included the phosphotransferase system and the three major allosteric regulations under physiological conditions. Moreover, we did not consider feedback regulation by ATP, ADP, or AMP that were shown to cause oscillations in earlier models[15,16].

Our results indicated that high saturation and high cooperativity of PDH promoted pyruvate oscillations. If the enzyme is already saturated by its substrate, its buffering capacity is limited, and the reaction rate can no longer be sensitively modulated by its substrate concentration. Recent studies have shown that enzymes with saturation kinetics can buffer substrate concentration changes[3,21,22]. Moreover, it was reported

that reactions at branch points out of autocatalytic cycles, like PDH in this case, need to have weak substrate affinities (low saturation) to ensure stability of the autocatalytic cycle[23]. This is supported by literate binding constants and pyruvate levels that indicate overabundance of PDH[24,25]. Thus, in practice, other factors are presumably more likely to cause pyruvate oscillations.

Positive feedforward regulations amplify signals and the feedforward activation of the PYK by FBP was shown to increase robustness of glycolysis[14]. Yet, our results indicated that, given the right circumstances, the feedforward activation can also promote sustained oscillations of pyruvate concentrations. This behaviour agrees with the theory of highly optimized tolerance, which states that robustness (tolerance) against general perturbations can be accompanied by fragilities against specific perturbations[15,26].

Here, we showed that regulation of glycolysis and enzyme saturation can cause sustained oscillations in pyruvate levels with an average period of three minutes. We show that oscillations can occur, even if feedback regulation by ATP, ADP, and AMP are not considered. It will be important to clarify the causes of the oscillations in more detail, for example by using a more comprehensive mechanistic model, and experimental time-resolved measurements of pyruvate levels.

5.5 Methods

The stoichiometry of the model is shown in **Figure 30**. Mass balancing results in a system of ordinary differential equations (ODEs), F, which is a temporal function of the state variables x (G6P, FBP, PEP, PYR) and the kinetic parameters p. In total, the system comprises four variables and 21 kinetic parameters. Dilution of metabolites by growth was not considered due to large differences between growth dilution and glycolytic flux.

$$F(x,p) = \frac{dx}{dt} = \begin{cases} PTS + FBPase - PFK \\ PFK - FBPase - FBA \\ 2 \cdot FBA - PYK - PTS \\ PTS + PYK - PDH \end{cases} \quad (5.1)$$

The six reactions (PTS, PFK, FBPase, FBA, PYK, PDH) are described by the following kinetic equations:

The PTS reaction takes up glucose from outside the system boundary depending on the ratio of PYR/PEP. With an assumed glucose uptake rate of 8 mmol g^{-1} h^{-1} and the specific cell volume for *E. coli* (2 µl mg^{-1}) the reaction rate for the PTS system is[27]:

$$r1 = PTS = \frac{Vmax}{k1 \cdot \frac{PYR}{PEP} + k2 + k3 \cdot \frac{PYR}{PEP} + 1} = \frac{8\ mmol\ g^{-1}h^{-1}}{0.002\ l\ g^{-1}} \cdot \frac{1h}{60\ min} = 66.66 \frac{mM}{min} \quad (5.2)$$

Reaction 5.2 (PFK) follows hill-type kinetics as it was shown that the enzyme exhibits cooperative kinetics towards its substrate[28]. The enzyme is allosterically inhibited by PEP which is modelled by a power-law where n1 is the hill-coefficient and a1 is the power-law exponent.

$$r2 = PFK = \frac{Vmax2}{\left(1 + \frac{Km1}{G6P}\right)^{n1}} \cdot PEP^{-a1} \quad (5.3)$$

Reaction 5.3 (FBPase) is modelled by a Michaelis-Menten type kinetic. The activation of FBPase by PEP is modelled by a power-law term.

$$r3 = FBPase = Vmax3 \cdot \frac{FBP}{FBP + Km2} \cdot PEP^{a2} \quad (5.4)$$

The flux ratio between PFK and FBP is randomly sampled between 0.01 and 1 and constraint to a net flux of 66.66 mmol l^{-1} min^{-1}. **Reaction 5.4** (FBA) is modelled by a Michaelis Menten type kinetic. Here, glycolysis is simplified by lumping four reactions into one reaction (FBA-ENO):

$$r4 = FBA = V_{max4} \cdot \frac{FBP}{FBP + Km3} \quad (5.5)$$

Reaction 5.5 (PYK) follows hill kinetics. The allosteric feed forward activation by FBP is modelled by a power-law[9]:

$$r5 = PYK = \frac{Vmax5}{\left(1 + \frac{Km4}{PEP}\right)^{n2}} \cdot FBP^{a3} \quad (5.6)$$

Reaction 5.6 (PDH) follows hill kinetics:

$$r6 = PDH = \frac{Vmax6}{\left(1 + \frac{Km5}{PYR}\right)^{n3}} \quad (5.7)$$

The 21 kinetic parameters were randomly sampled 2,000 times from a log-uniform distribution. All state variables were set to a concentration of 1 mM. The system was set

to a steady state by first setting the total reaction flux of all net reactions to the glucose uptake rate. Then, parameter values of binding constants (*k1-k3*, K_{m1}-K_{m5}), hill coefficients (*n*1-*n*3) and power-law exponents (exponents *α*1-*α*3) were inserted. Then the maximum velocities were calculated (V_{max1}-V_{max6}). Binding constants were sampled between 0.1 mM and 10 mM, hill-coefficients were sampled between 1 and 4, where a coefficient of 1 resembles a Michaelis-Menten type kinetic, and power-law exponents for allosteric regulations were sampled between 1 and 4 for allosteric activations and -1 and -4 for feedback inhibitions. The perturbation of the glucose uptake rate was simulated by changing the uptake rate by 5% (or 10-fold) at 50 minutes simulation time. The resulting time-course data were then processed to identify parameter sets which led to oscillations. First, the time courses were fitted with a first order polynomial to remove trends and to align the time courses. Second, the data were Fourier transformed from time domain to frequency domain. The MATLAB function “findpeaks” was then used to find signals with amplitudes above 0.01. The corresponding time courses and parameter sets were then selected.

5.6 References

(1) Kitano, H. Biological Robustness. *Nat Rev Genet* **2004**, *5* (11), 826–837. https://doi.org/10.1038/nrg1471.

(2) Reaves, M. L.; Young, B. D.; Hosios, A. M.; Xu, Y.-F.; Rabinowitz, J. D. Pyrimidine Homeostasis Is Accomplished by Directed Overflow Metabolism. *Nature* **2013**, *500* (7461), 237–241. https://doi.org/10.1038/nature12445.

(3) Sander, T.; Farke, N.; Diehl, C.; Kuntz, M.; Glatter, T.; Link, H. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. Coli by Enforcing Enzyme Overabundance. *Cell Syst* **2019**, *8* (1), 66-75.e8. https://doi.org/10.1016/j.cels.2018.12.005.

(4) Grimbs, S.; Selbig, J.; Bulik, S.; Holzhütter, H.-G.; Steuer, R. The Stability and Robustness of Metabolic States: Identifying Stabilizing Sites in Metabolic Networks. *Mol Syst Biol* **2007**, *3*, 146. https://doi.org/10.1038/msb4100186.

(5) Tran, L. M.; Rizk, M. L.; Liao, J. C. Ensemble Modeling of Metabolic Networks. *Biophys J* **2008**, *95* (12), 5606–5617. https://doi.org/10.1529/biophysj.108.135442.

(6) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat Biotechnol* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.

(7) Steuer, R.; Gross, T.; Selbig, J.; Blasius, B. Structural Kinetic Modeling of Metabolic Networks. *Proc Natl Acad Sci U S A* **2006**, *103* (32), 11868–11873. https://doi.org/10.1073/pnas.0600013103.

(8) Chubukov, V.; Gerosa, L.; Kochanowski, K.; Sauer, U. Coordination of Microbial Metabolism. *Nat Rev Microbiol* **2014**, *12* (5), 327–340. https://doi.org/10.1038/nrmicro3238.
(9) Link, H.; Kochanowski, K.; Sauer, U. Systematic Identification of Allosteric Protein-Metabolite Interactions That Control Enzyme Activity in Vivo. *Nat Biotechnol* **2013**, *31* (4), 357–361. https://doi.org/10.1038/nbt.2489.
(10) Bogorad, I. W.; Lin, T.-S.; Liao, J. C. Synthetic Non-Oxidative Glycolysis Enables Complete Carbon Conservation. *Nature* **2013**, *502* (7473), 693–697. https://doi.org/10.1038/nature12575.
(11) Wang, C.-Y.; Lempp, M.; Farke, N.; Donati, S.; Glatter, T.; Link, H. Metabolome and Proteome Analyses Reveal Transcriptional Misregulation in Glycolysis of Engineered E. Coli. *Nat Commun* **2021**, *12*, 4929. https://doi.org/10.1038/s41467-021-25142-0.
(12) Marín-Hernández, A.; Gallardo-Pérez, J. C.; Rodríguez-Enríquez, S.; Encalada, R.; Moreno-Sánchez, R.; Saavedra, E. Modeling Cancer Glycolysis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2011**, *1807* (6), 755–767. https://doi.org/10.1016/j.bbabio.2010.11.006.
(13) Sel'Kov, E. E. Self-Oscillations in Glycolysis. 1. A Simple Kinetic Model. *Eur J Biochem* **1968**, *4* (1), 79–86. https://doi.org/10.1111/j.1432-1033.1968.tb00175.x.
(14) Kremling, A.; Bettenbrock, K.; Gilles, E. D. A Feed-Forward Loop Guarantees Robust Behavior in Escherichia Coli Carbohydrate Uptake. *Bioinformatics* **2008**, *24* (5), 704–710. https://doi.org/10.1093/bioinformatics/btn010.
(15) Chandra, F. A.; Buzi, G.; Doyle, J. C. Glycolytic Oscillations and Limits on Robust Efficiency. *Science* **2011**, *333* (6039), 187–192. https://doi.org/10.1126/science.1200705.
(16) Eschrich, K.; Schellenberger, W.; Hofmann, E. Oscillations in the Phosphofructokinase--Fructose 1,6-Bisphosphatase Cycle. II. Influence of Fructose 1,6-Bisphosphatase on the Character of Oscillatory States. *Biomed Biochim Acta* **1983**, *42* (6), 609–621.
(17) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; Feist, A. M.; Palsson, B. O. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat Biotechnol* **2017**, *35* (10), 904–908. https://doi.org/10.1038/nbt.3956.
(18) Kochanowski, K.; Volkmer, B.; Gerosa, L.; Haverkorn van Rijsewijk, B. R.; Schmidt, A.; Heinemann, M. Functioning of a Metabolic Flux Sensor in Escherichia Coli. *Proc Natl Acad Sci U S A* **2013**, *110* (3), 1130–1135. https://doi.org/10.1073/pnas.1202582110.
(19) Schink, S. J.; Christodoulou, D.; Mukherjee, A.; Athaide, E.; Brunner, V.; Fuhrer, T.; Bradshaw, G. A.; Sauer, U.; Basan, M. Glycolysis/Gluconeogenesis Specialization in Microbes Is Driven by Biochemical Constraints of Flux Sensing. *Molecular Systems Biology* **2022**, *18* (1). https://doi.org/10.15252/msb.202110704.
(20) Smallbone, K.; Messiha, H. L.; Carroll, K. M.; Winder, C. L.; Malys, N.; Dunn, W. B.; Murabito, E.; Swainston, N.; Dada, J. O.; Khan, F.; Pir, P.; Simeonidis, E.; Spasić, I.; Wishart, J.; Weichart, D.; Hayes, N. W.; Jameson, D.; Broomhead, D. S.; Oliver, S. G.; Gaskell, S. J.; McCarthy, J. E. G.; Paton, N. W.; Westerhoff, H. V.; Kell, D. B.; Mendes, P. A Model of Yeast Glycolysis Based on a Consistent Kinetic

Characterisation of All Its Enzymes. *FEBS Lett* **2013**, *587* (17), 2832–2841. https://doi.org/10.1016/j.febslet.2013.06.043.

(21) Fendt, S.-M.; Buescher, J. M.; Rudroff, F.; Picotti, P.; Zamboni, N.; Sauer, U. Tradeoff between Enzyme and Metabolite Efficiency Maintains Metabolic Homeostasis upon Perturbations in Enzyme Capacity. *Mol Syst Biol* **2010**, *6*, 356. https://doi.org/10.1038/msb.2010.11.

(22) Davidi, D.; Milo, R. Lessons on Enzyme Kinetics from Quantitative Proteomics. *Current Opinion in Biotechnology* **2017**, *46*, 81–89. https://doi.org/10.1016/j.copbio.2017.02.007.

(23) Barenholz, U.; Davidi, D.; Reznik, E.; Bar-On, Y.; Antonovsky, N.; Noor, E.; Milo, R. Design Principles of Autocatalytic Cycles Constrain Enzyme Kinetics and Force Low Substrate Saturation at Flux Branch Points. *eLife* **2017**, *6*, e20667. https://doi.org/10.7554/eLife.20667.

(24) Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Res* **2002**, *30* (1), 47–49. https://doi.org/10.1093/nar/30.1.47.

(25) Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers—the Database of Key Numbers in Molecular and Cell Biology. *Nucleic Acids Res* **2010**, *38* (Database issue), D750–D753. https://doi.org/10.1093/nar/gkp889.

(26) Carlson, J. M.; Doyle, J. Highly Optimized Tolerance: Robustness and Design in Complex Systems. *Phys. Rev. Lett.* **2000**, *84* (11), 2529–2532. https://doi.org/10.1103/PhysRevLett.84.2529.

(27) Lee, Y.; Lafontaine Rivera, J. G.; Liao, J. C. Ensemble Modeling for Robustness Analysis in Engineering Non-Native Metabolic Pathways. *Metab Eng* **2014**, *25*, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.

(28) Evans, P. R.; Farrants, G. W.; Hudson, P. J. Phosphofructokinase: Structure and Control. *Philos Trans R Soc Lond B Biol Sci* **1981**, *293* (1063), 53–62. https://doi.org/10.1098/rstb.1981.0059.

Chapter 6 Mapping regulation in *E. coli* metabolism with a metabolism wide CRISPRi library

Alejandra Alvarado, Andreas Verhülsdonk, Niklas Farke, Johanna Rapp, Sevvalli Thavapalan, Felicia Troßmann, Timo Glatter, Hannes Link

Lempp, M.; Farke, N.; Kuntz, M.; Freibert, S. A.; Lill, R.; Link, H. Systematic Identification of Metabolites Controlling Gene Expression in E. Coli. Nat Commun 2019, 10 (1), 4463. https://doi.org/10.1038/s41467-019-12474-1.

Parts of this chapter are published in Nature Communications 2019, 10 (1), 4463. My contribution to this work included the literature search for known metabolite-Transcription factor interactions and the graph-based distance analysis.

6.1 Abstract

Cells must regulate enzyme levels because low amounts of enzymes limit metabolic flux and high amounts of enzymes are costly. Enzyme levels are regulated by transcriptional regulation, and metabolites are the key signalling molecules. Here, we used CRISPR interference (CRISPRi) to knockdown gene expression of 283 metabolic genes in *E. coli*. The measured proteome data showed compensatory changes to offset the knockdown and indicated transcriptional feedback loops in 24 metabolic pathways. Untargeted metabolomics showed that the perturbation offsets the metabolome, as indicated by substrate accumulations of 66 CRISPRi target genes. Integrating proteome and metabolome data then allowed us to recover known metabolite-TF interactions, as well as suggest nicotinate as a putative metabolite effector for the transcription factor NadR. In conclusion, by integrating proteomics and metabolomics data from 283 CRISPRi knockdowns, we could map feedback regulation in *E. coli* metabolism.

6.2 Introduction

Microbes constantly face internal and external perturbations. To withstand these perturbations, they have evolved mechanisms that ensure robustness[1,2]. One of the most important mechanisms that allows cells to adapt is dynamic feedback control that is executed via functional metabolite-protein interactions. This way, metabolites can allosterically regulate the activity of metabolic enzymes, and by means of metabolite-transcription factor interactions they can control enzyme abundances[3–5]. To fully understand *E. coli*, we first need to identify all regulatory interactions. Although many allosteric interactions are known[6], a recent study identified a hitherto unprecedented number of metabolite-protein interactions, indicating that we are still unaware of many functional interactions[7].

Historically, metabolite-protein interactions were identified by *in-vitro* assays that combined single proteins with single metabolites[8,9]. Recently systematic approaches based on mass spectrometry emerged to study these interactions at high throughput and at scale. Usually, they either focus on protein changes or on metabolite changes that occur during the interaction[10]. Mass spectrometry-based approaches that focus on

conformational protein changes during metabolite-protein interactions include for example approaches based on limited proteolysis, or thermal proteome profiling[7,11,12]. Conversely, a metabolite-centric approach used nuclear magnetic resonance spectroscopy to identify several metabolites that interact with specific proteins[13]. While these approaches enable to systematically identify metabolite-protein interactions, an interaction alone does not imply an underlying function. To infer interactions with physiological functions, multi-omics data are usually integrated. This was done using a multi-omics data set comprising 25 growth conditions in yeast. Paired with prior knowledge in form of simple mathematical models, functional allosteric interactions were identified[14]. Commonly, functional interactions are identified by observing metabolite or protein changes that compensate perturbations such as gene knockouts[15]. Gene knockouts, however, cause strong physiological changes and may hide subtle regulatory effects. To identify subtle regulation, knockdowns executed by CRISPR interference (CRISPRi) represent an alternative approach[16]. For example, CRISPRi has been used in *E. coli* to identify the mode of action of antimicrobial agents[17].

In this chapter, we set out to map functional metabolite-TF interactions in *E. coli*. Therefore, we used CRISPRi to knockdown gene expression of 283 metabolic genes that are essential for growth on glucose minimal medium[18]. The proteome data showed compensatory changes to offset the knockdowns, and a pathway analysis revealed transcriptional feedback loops in 24 metabolic KEGG pathways. Untargeted metabolomics showed substrate accumulations for 66 CRISPRi targets, indicating that the knockdowns offset the metabolome. Integrating proteome and metabolome data identified known metabolite-TF interactions in arginine and methionine biosynthesis, as well as a novel putative feedback loop between nicotinate and NadR in NAD *de novo* biosynthesis. In conclusion, we show that CRISPRi knockdowns of metabolic genes allows us to systematically map transcriptional feedback regulation in *E. coli* metabolism.

6.3 CRISPRi knockdowns cause downregulation of target genes

To identify feedback loops, we perturbed 283 genes in primary metabolism of *E. coli* with CRISPRi. Target genes with growth phenotypes were selected based on a previous

screen and were distributed across primary metabolism[18]. We measured the proteome and metabolome of each *E. coli* knockdown strain after inducing CRISPRi for 6.5 h. We first inspected the proteome data and observed that CRISPRi blocks transcription of target genes, causing dilution of the corresponding enzymes by growth. Therefore, we expected that the enzyme product of the targeted gene was downregulated. In 128/283 (45 %) strains, we observed a significant perturbation of the target enzyme (z-score < -3), indicating a functional knockdown (**Figure 34a**). 216/283 (76 %) of the target genes had at least a z-score < -1. The genes *aroD, plsC, lptE, coaA, coaD, lpxK, dapE, dcd, bioB* and *ispU* showed the strongest downregulation with an average z-score of -6.8. Against our expectations, two target genes were upregulated, which could be explained by transcriptional crosstalk, or a non-functional perturbation caused by errors in the sgRNA plasmid or dcas9 mutation. This is supported by an average of 40 significantly up- and downregulated genes for each knockdown (z-score < -3 or z-score > 3). The target genes *spoT, cysD, cysN, murF, metE, accB, lpxC, cysU* and *cysJ* had the highest number of proteome changes, indicating an influence of global, growth rate dependent regulation[19–21] (**Figure 34b**).

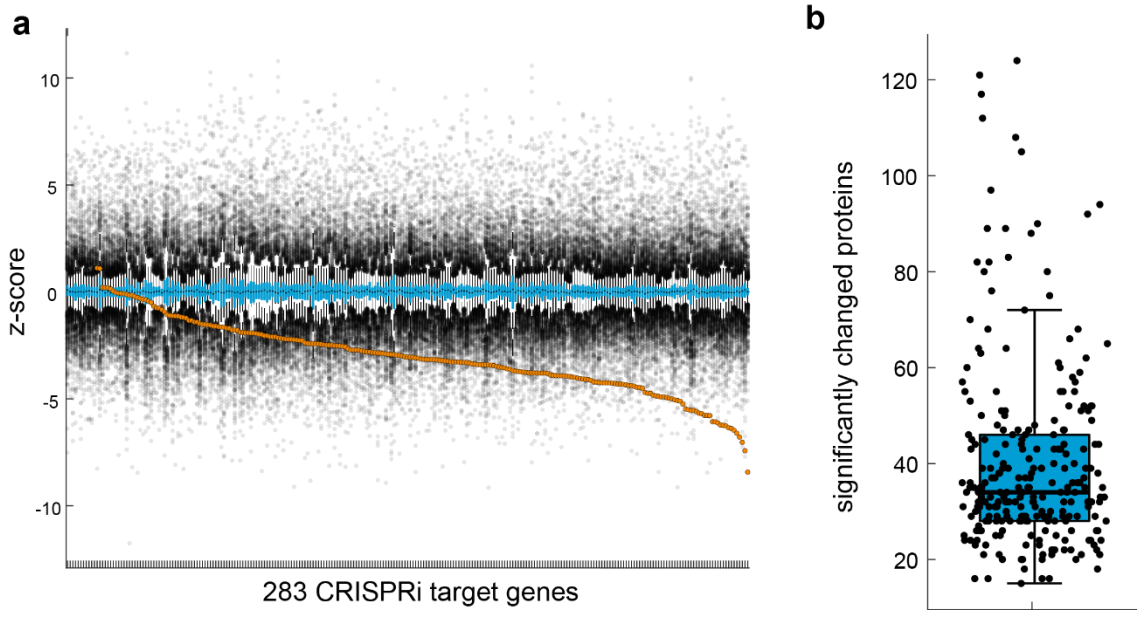


Figure 34. Knockdown of target genes causes proteome changes. a) *Boxplot and overlayed scatter plot. Each column is a CRISPRi target gene. Boxplots are shown in blue and whiskers are black. Upper and lower box edges indicate the 25 % and 75 % percentiles. The whiskers indicate the furthest point, at which samples were not considered as outliers. The black line indicates the median. Black dots are outliers. Orange dots are the CRISPRi targets.* ***b)*** *Boxplot showing the number of significantly changed proteins (z-score < -3 or z-score > 3) for each CRISPRi knockdown. Black dots are CRISPRi strains. Upper and lower box edges indicate the 25 % and 75 % percentiles. The whiskers indicate the furthest point, at which samples were not considered as outliers. The black line indicates the median.*

6.4 Proteome changes reveal feedback loops in 24 metabolic pathways

After having confirmed that the CRISPRi knockdowns offset the proteome, we next analysed the proteome data for compensatory feedback mechanisms. Since transcriptional feedback regulation is often pathway specific, we decided to take a pathway-based approach. Therefore, we assigned 43 KEGG (Kyoto Encyclopedia of Genes and Genomes)[22] pathway identifier to the target genes and each measured protein. For 24 KEGG pathways, we found at least one compensatory upregulation of genes (z-score > 2) within the same KEGG pathway (**Figure 35a**), strongly indicating compensatory feedback regulation. This is supported by downregulation of the respective CRISPRi targets in those pathways (on average 3-fold), except for *gpmA* and *ppc* which were upregulated (**Figure 35b**).

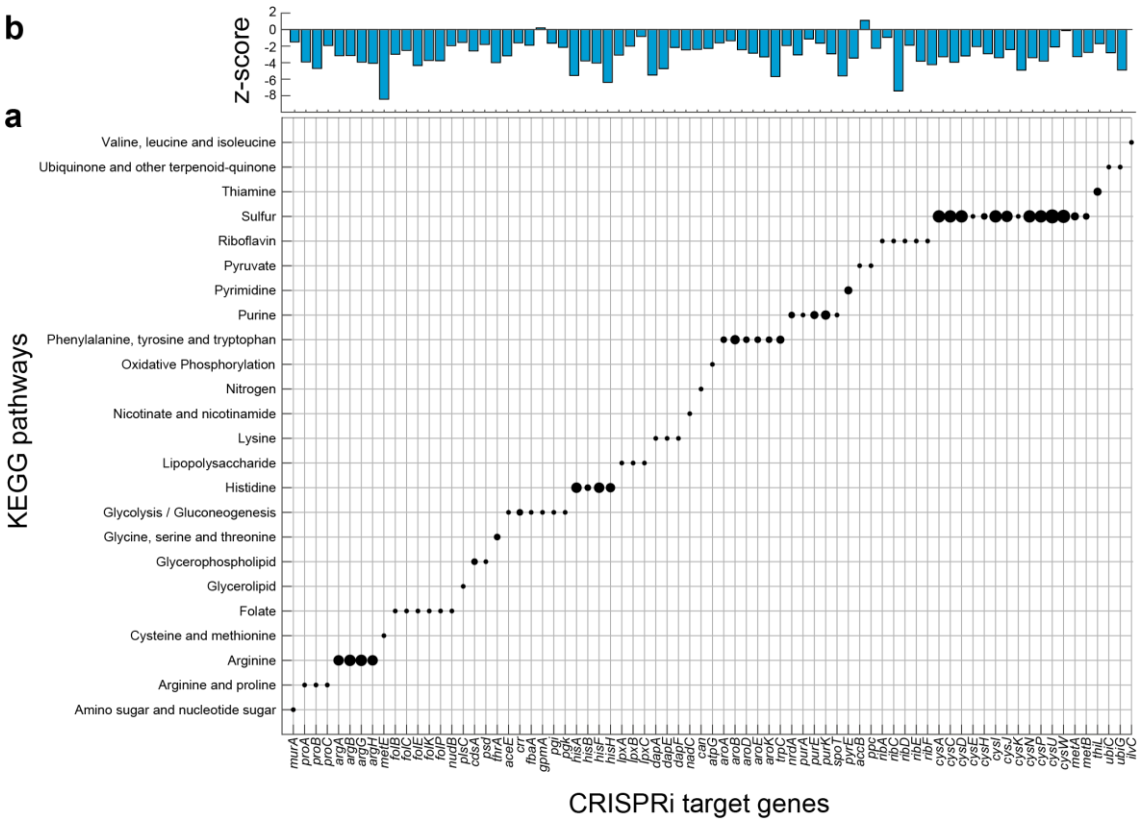


Figure 35. KEGG pathways with putative feedback circuits. a) *Target genes were assigned to one KEGG pathway. A dot indicates that the knockdown of a target gene causes an upregulation of at least one other target gene within the same KEGG pathway. Dot size corresponds to the number of genes in the KEGG pathway that are upregulated above a z-score of two.* ***b)*** *z-scores of the proteins corresponding to the perturbed target genes.*

We quantified the strength of the compensatory response by counting the number of pathway genes that were upregulated and had at least a z-score above two. The pathways with the strongest responses were arginine biosynthesis, histidine biosynthesis and sulfur metabolism. For arginine biosynthesis it is known that arginine levels control expression of the pathway genes by modulating the activity of the transcription factor ArgR. Sulfur metabolism is controlled via CysB that is regulated by O-acetyl-serine. Similarly, histidine controls expression of histidine genes via transcriptional attenuation. For example, the target genes *hisA/B/C/D/F/H* are all part of histidine biosynthesis and together with *hisG/I* were assigned to the histidine KEGG pathway.

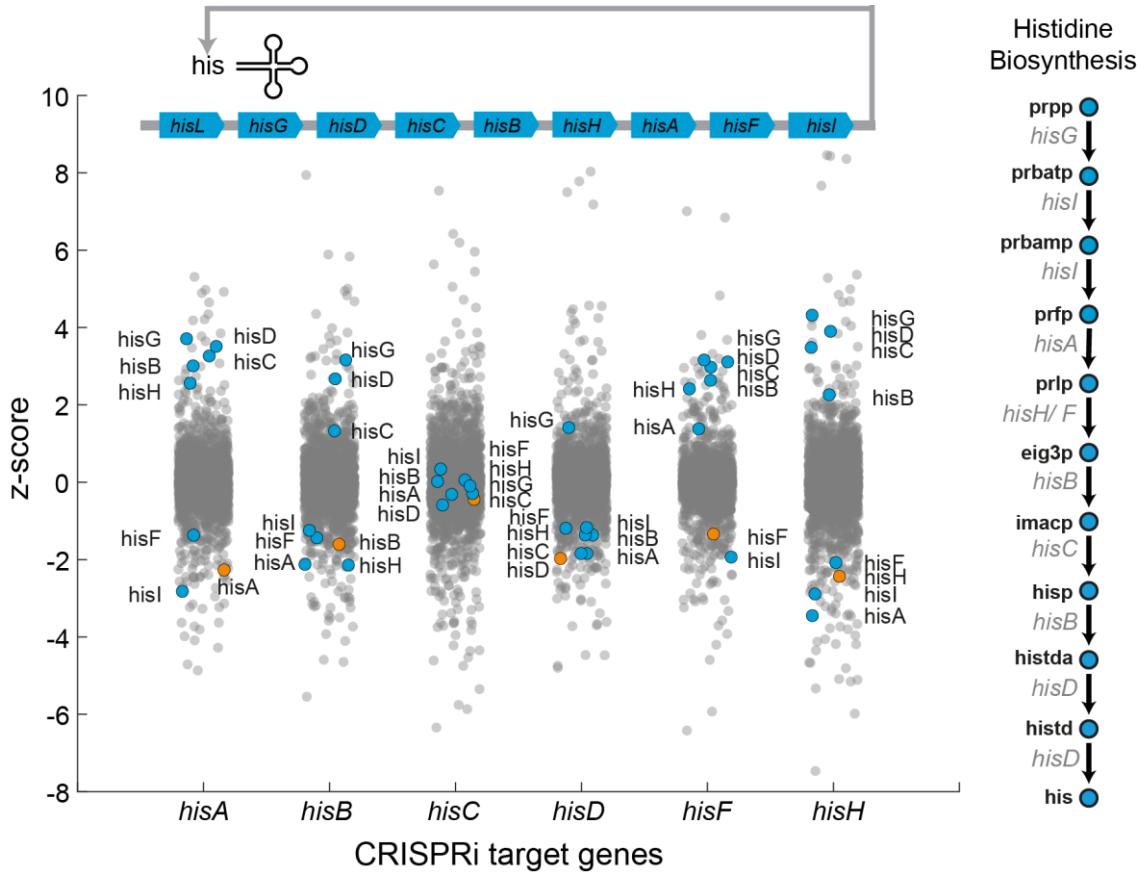


Figure 36*. *Histidine KEGG pathway. *Dots are all measured proteins. Blue dots are measured genes in the histidine KEGG pathway. Orange dots are CRISPRi target genes. Above the scatter plot, the operon structure of the histidine pathway is shown. The loops indicate transcriptional attenuation of the pathway. On the right the metabolic histidine pathway is shown. Blue dots are metabolites. Arrows are metabolic reactions.*

Except for *hisC*, we observed a downregulation of the target genes. However, we also observed that genes downstream of the target gene were downregulated. For example,

perturbation of *hisA* causes downregulation of *hisF* and *hisI*. Similarly, a perturbation of *hisD* led to a downregulation of all genes except for *hisG*. This polar effect is caused by the operon structure. CRISPRi blocks transcription within the operon. Thus, all genes located upstream of the CRISPRi target are still getting transcribed by the RNA polymerase, while transcription of all downstream targets is blocked (**Figure 36**). While the genes downstream of the target gene were downregulation, genes upstream of the target gene were upregulated, indicating a compensatory feedback loop. Thus, we identified feedback regulation in 24 metabolic KEGG pathways. Besides identifying pathways that are feedback regulated, we also identified pathways that are likely not feedback regulated under these conditions. For example, we perturbed seven targets in the methylerythritol phosphate (MEP) pathway (**Figure 37**). In all cases, the target genes were downregulated, but none of the pathway enzymes were upregulated. In the knockdown of *ispD*, *ispF* was also downregulated because it is located downstream in the same operon. This indicated that the MEP pathway is not controlled by transcriptional regulation.

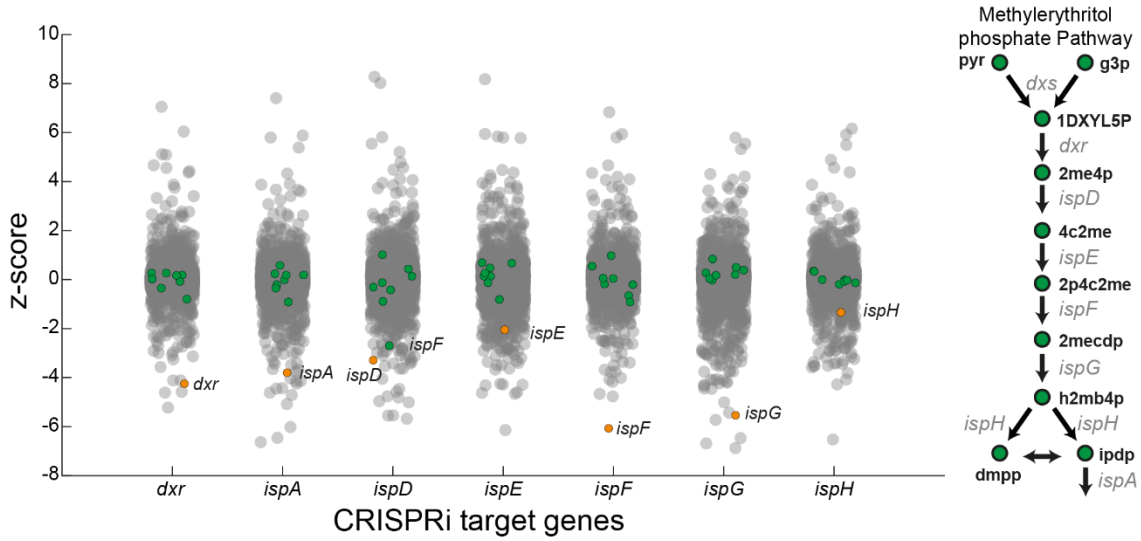


***Figure 37. Methylerythritol phosphate KEGG pathway**. Dots are all measured proteins. Green dots are measured genes in the methylerythritol phosphate kegg pathway. Orange dots are CRISPRi target genes. On the right, the metabolic histidine pathway is shown. Green dots are metabolites. Arrows are metabolic reactions.*

Moreover, the pathway analysis indicated transcriptional regulation for several genes, although there is no transcriptional regulation known. For example, the perturbations of *ubiC/E/G* caused an upregulation of *menH*. However, the operon of *menH* is not

transcriptionally regulated. Likewise, knockdown of *folB/C/E/K/P* and *nudB* caused a clear accumulation of *folM*, and the operon of *folM* is not known to be transcriptionally regulated (ArgR was predicted to control the operon that contains *folM*, but it was not confirmed yet). Knockdown of *folA* caused a downregulation of folB although both genes are not located in the same operon, indicating that a metabolite close to *folA* could be the regulator. The genes *folA* and *folM* code for different hydrofolate reductases and are involved in the final step of the tetrahydrofolate biosynthesis. It is surprising that *folM* is downregulated when we perturb *folA*, since the data indicate a compensatory feedback mechanism. This indicates that a transcriptional feedback mechanism is unknown for the *folM* operon (**Figure 38**).

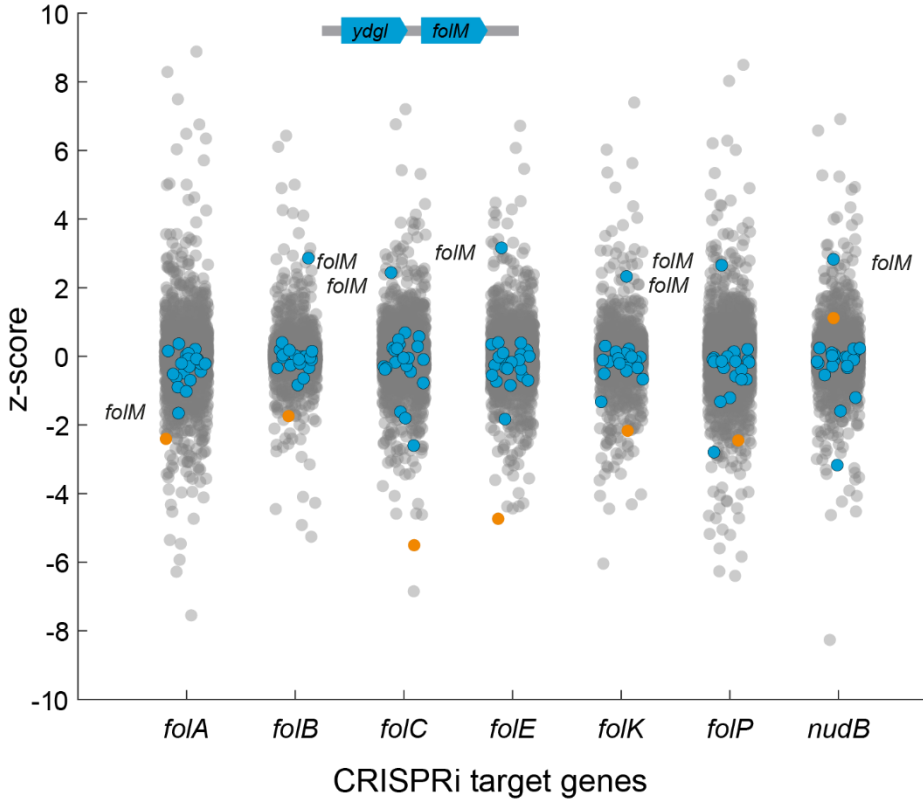


Figure 38. Folate metabolism*. Dots are all measured proteins. Blue dots are measured genes in the methylerythritol phosphate KEGG pathway. Orange dots are CRISPRi target genes shown on the x-axis. On top the operon structure of the folM gene is shown.*

In conclusion, a pathway-based approach identified feedback regulation in 24 KEGG pathways using proteome data from CRISPRi knockdowns. Moreover, we identified likely knowledge gaps in the TRN of *E. coli* in the operons of *folM*, and *menH*. We showed that polar effects make it more difficult to identify feedback circuits, because they cause additional downregulations, and likely also additional metabolite changes. While gene

knockdowns cause compensatory responses by gene upregulations, we can't exclude the possibility that, in some rare cases, downregulations can also compensate the knockdowns by other mechanisms. Thus, it will be important to include operon structure in the systematic search for feedback circuits. Next, we went into literature to find known metabolite-TF interactions to guide our search for new functional metabolite-TF interactions.

6.5 Transcriptional regulation in proximity of the effector metabolite

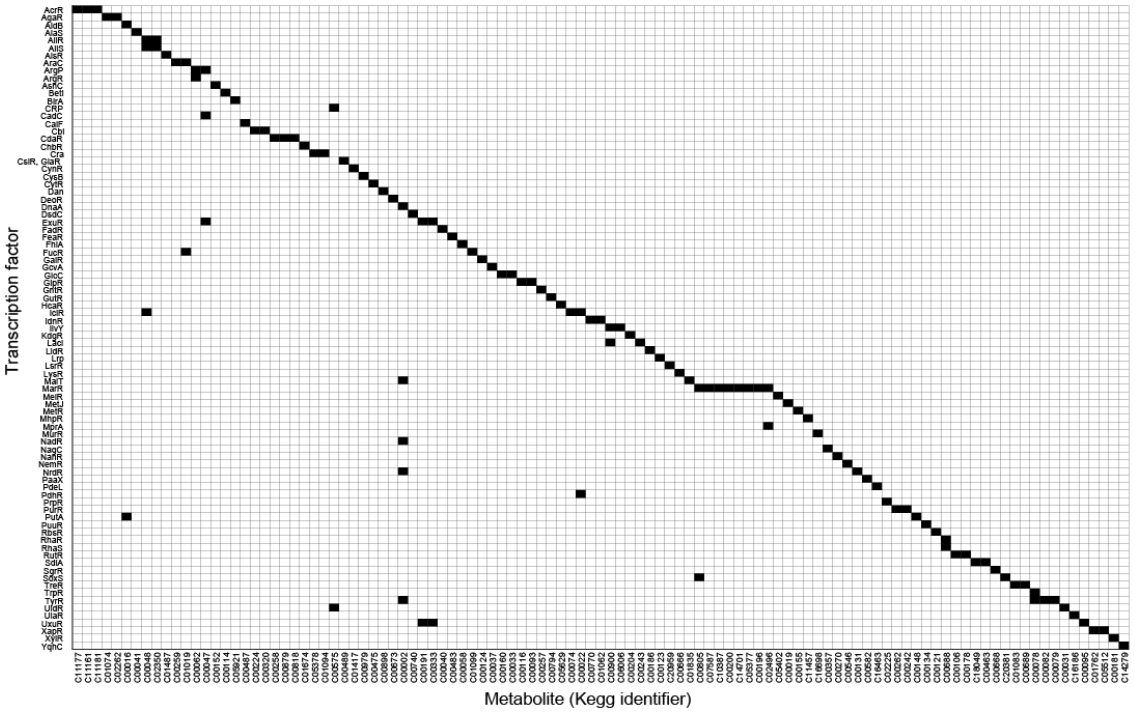


***Figure 39**. **Metabolite-TF interaction map**. Metabolite-transcription factor interactions that are described in the literature and databases. Shown are 87 transcription factors (rows) and their respective effector metabolites (columns). Black boxes are known metabolite-transcription factors interactions. This figure is published in Nat Commun 2019, 10 (1), 4463[23].*

Next, we mapped 134 unique metabolite-TF interactions for *E. coli* (**Figure 39**) that involved 87 unique TFs based on entries in (RegulonDB[24]). 80 % of the interacting metabolites were unique, indicating specific metabolism-wide crosstalk between metabolism and transcription. The remaining metabolite were mostly promiscuous effectors, based on structural similarity to the main effector. For example, the tyrosine repressor TyrR has an affinity towards L-tyrosine, L-tryptophan, and L-phenylalanine, which is likely due the common aromatic structure. The metabolite effectors could be

classified into different chemical groups: amino acids, nucleobase derivatives, organic acids, sugars, and others. Out of those classes, organic acids were the most frequent effectors (**Figure 40a**). We then wondered if transcription factors are regulated by proximal metabolites, or by metabolites that are further away in the metabolic network. To calculate distances of metabolite-TF interactions, we used the iJO1366 genome scale model. First, we removed all cofactors and non-intracellular metabolites of the stoichiometric matrix of the mode.

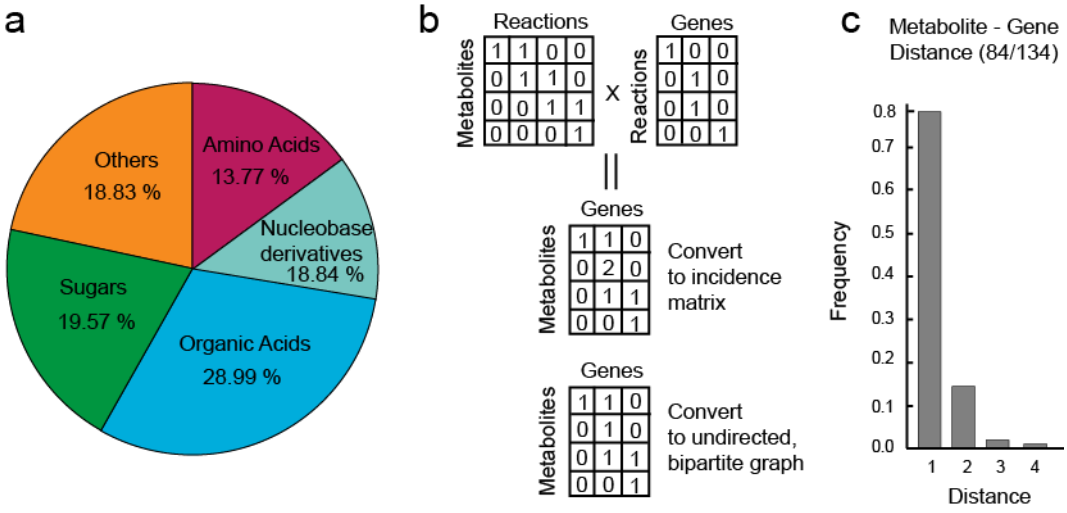


Figure 40. a) Classification of effector metabolites into chemical groups. b) *Scheme explaining how the bipartite metabolite-gene network is obtained from the stoichiometric matrix and the reaction-gene matrix of the genome scale model iJO1366.* ***c)*** *Distance between a metabolite and the target ganes of the interacting transcription factor. The distance d was transformed by the following equation to account only for genes: Distance = (d+1)/2. Parts of this figure are published in Nat Commun 2019, 10 (1), 4463.*

Then, we created a metabolite-gene adjacency matrix, F, by calculating the inner product of the modified stoichiometric matrix, N, and the reaction gene matrix, G. Next, we calculated the Boolean of F, F'. We then transformed F' into an undirected, bipartite graph, in which nodes are metabolites and genes. We then used the bipartite graph to calculate the metabolite-gene distances between all pairs of nodes and stored the information in the distance matrix, D (**Figure 40b**). For known metabolite-TF interactions, we calculated the distances between the regulating metabolite and each of the target genes of the respective TF. Then, we took the smallest distance. For metabolites that are not part of the model, we omitted the distance calculation, leaving 84 metabolite-gene distances[6]. Strikingly, 80 % of the 84 metabolite-gene interactions were directly adjacent, indicating that regulatory metabolites modulate expression of

proximal genes (**Figure 40c**). Here, we showed that TFs are usually regulated by metabolites in proximity of target genes. Next, we sought to identify effector metabolites that modulate changes in TF activity.

6.6 CRISPRi knockdowns cause substrate accumulations

To screen for effector metabolites that cause the proteome changes in the targeted KEGG pathways, we first measured the metabolome of 283 CRISPRi knockdowns with flow-injection metabolomics (FI-MS) (n = 1). We annotated 370 and 528 ion peaks to metabolites in positive and negative mode, respectively. To test if the metabolome responds in a similar way as the proteome, we counted the number of strongly (z-score < -2 and z-score > 2) changed ion peaks that were annotated to single (de-)protonated metabolites. On average, six and seven metabolites changed per strain in negative and positive mode, respectively. Combining positive and negative ionization mode, 11.2 metabolites changed on average (**Figure 41**), indicating that the CRISPRi knockdowns offset the metabolome.

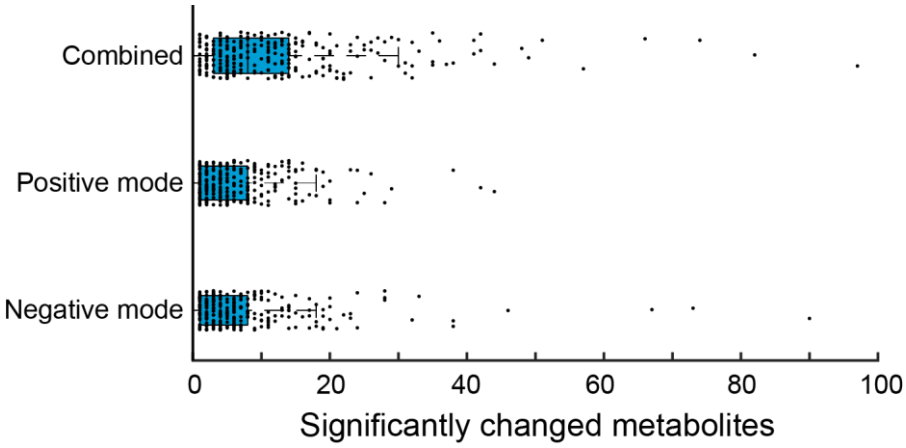


***Figure 41**. **CRISPRi knockdowns offset the metabolome**. Boxplot of metabolome changes in negative mode, positive mode, and in both modes combined. Upper and lower box edges indicate the 25 % and 75 % percentiles. The whiskers indicate the furthest point, at which samples were not considered as outliers. The black line indicates the median. Black dots are CRISPRi strains.*

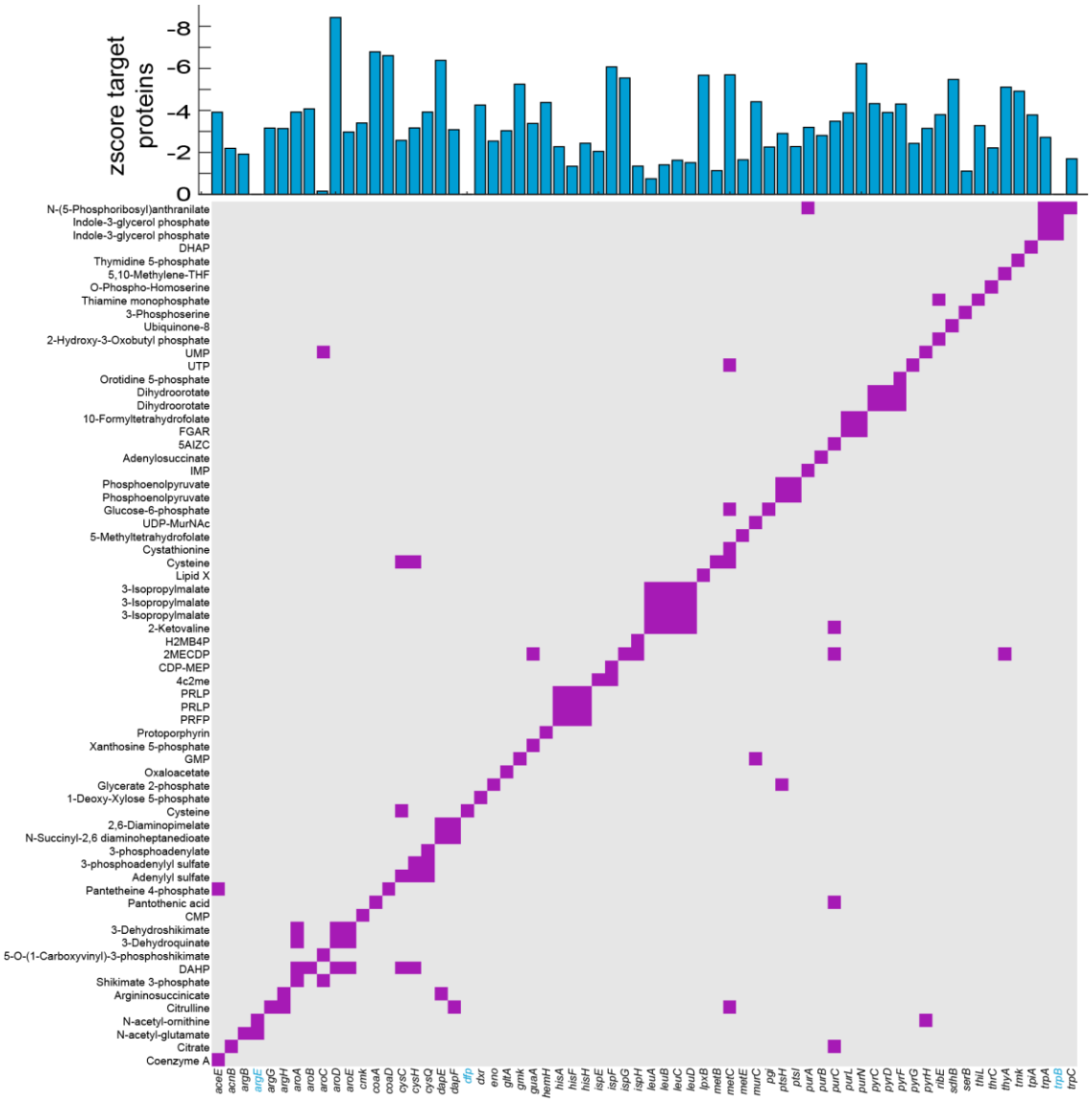


Figure 42. CRISPRi knockdowns cause Substrate accumulations. *Heatmap with CRISPRi target genes on the x-axis, and substrate metabolites on the y-axis. Target gene-substrate pairs are on the main diagonal. A purple square indicates a zscore > 2. Blue gene names were not measured in the proteome data. The bar plot shows the zscore of the corresponding target genes in the proteome data.*

Next, we looked for accumulations of the substrates of the target genes. Whether a reactant of an enzyme is a substrate or a product is determined by the direction of the metabolic flux of a reaction. To estimate flux directions, we performed Flux Balance Analysis (FBA)[25] with the latest genome scale metabolic reconstruction iML1515[26,27]. Assuming growth on glucose minimal medium, we identified 427 reactions with non-zero flux. For reactions with negative flux, we exchanged substrates and products. For 203/283 target genes, we annotated at least one substrate metabolite. For 39/80 target genes without annotated substrate, FBA predicted that the corresponding reaction had

no flux. In total, we found substrate accumulations in 66 CRISPRi target strains (zscore > 2). In 51/66 cases, the perturbed target enzyme was downregulated, as expected (zscore < -2). In 10/15 remaining cases, the target enzymes were only slightly downregulated, since zscores were in the range between -1 and -2, and in 2/15 cases between 0 and -1. In three cases the target proteins were not measured (**Figure 42**).

In 83 cases, the CRISPRi targets were significantly downregulated, but the substrates did not accumulate. In 55/83 cases, the substrate had a small positive z-score (between 0 and 2), while in the remaining 28 cases the substrate had a small negative z-score (between 0 and -2). This indicated that for the majority of significantly changed target genes metabolite levels increased as expected, albeit not always significantly.

In conclusion, substrates accumulated in 66 CRISPRi target strains. While most target enzymes were strongly downregulated, in some cases substrates accumulated also for less downregulated target enzymes. A reason could be that the enzyme is fully saturated by its substrate and therefore cannot buffer the perturbation via enzyme activity. In 83 cases, however, we did not observe a substrate accumulation despite a strong knockdown of the target enzyme. There are many potential reasons why substrates did not accumulate. For example, ATP and ADP are the substrates of 28/83 CRISPRi targets. However, there are many metabolic reactions that interconvert ATP and ADP. This redundancy confers robustness and likely buffers changes in ATP and ADP levels. Therefore it seems unlikely to observe accumulations in ATP and ADP levels that are caused by single gene knockdowns.

6.7 Screening for metabolites that control gene expression using untargeted metabolomics

We next sought to identify metabolites that cause changes in TF activity. Transcription factors can activate or repress transcription, and metabolite binding can increase (activator) or decrease (inhibitor) TF activity. Metabolites can accumulate or deplete, and together with the metabolite mode (activator, inhibitor) and the TF mode (activator, repressor), there are in total eight possible modes how gene expression can be modulated.

To identify putative metabolites that control TF activity, we calculated the z-score for all annotated ion peaks. First, we verified known metabolite-TF interactions in the data. Therefore, we investigated the regulons of ArgR and MetR since our proteomics pathway analysis showed active feedback regulation in both pathways (**Figure 35**). Arginine is a known inhibitor of ArgR, and ArgR is a repressor of the arginine pathway. An upregulation of the arginine pathway is therefore achieved by arginine depletion. We observed strong arginine depletions in the CRISPRi strains *argA*, *argB*, *argG*, and *argH* (zscore <-2) and the proteomics data showed that these targets caused an upregulation of the arginine KEGG pathway (**Figure 43**). Similarly, a knockdown of *metE* caused a response of the cysteine and methionine KEGG pathway. Homocysteine is an activator of MetR, and homocysteine was among the highest upregulated metabolites, thus validating the proteome data (**Figure 43**). Both cases indicate that it should be possible to identify metabolites that control transcription with this dataset. However, arginine and homocysteine were strongly changed in many target genes, indicating that false positives could be problematic for the inference of functional metabolite-transcription factor interactions.

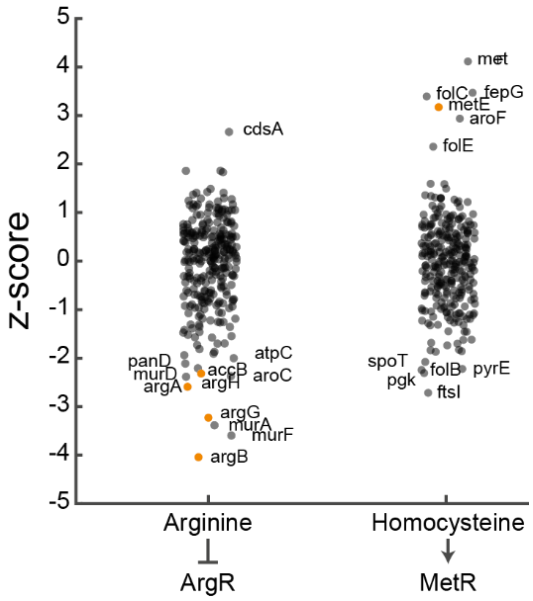


Figure 43. Arginine and Homocysteine changes in 283 CRISPRi knockdowns. a) *Y-axis shows the z-score of the metabolite in the respective strain. X-axis are known metabolite-TF interactions. Dots are 283 CRISPRi strains. Orange dots are CRISPRi strains with strong (zscore < -2, zscore > 2) changes in the respective metabolite (Arginine, Homocysteinel) that were likely to cause an upregulation of the corresponding KEGG pathway. CRISPRi strains with a zscore < -2 or a zscore > 2 are labelled. Orange dots are CRISPRi targets that caused an upregulation of the corresponding KEGG pathway.*

We next inspected the knockdown of *nadC* that caused a compensatory response in the nicotinate and nicotine amid KEGG pathway. The knockdown of *nadC* caused an upregulation of *nadA*, and *nadA* is regulated by the TF NadR. For NadR, however, only ATP is currently known as an effector metabolite[24] (**Figure 39**). The knockdown of *nadC* caused a strong and relatively specific (only the *dapB*, *sucB*, and *adk* knockdowns caused nicotinate accumulation (**Figure 44**)) accumulation of nicotinate (zscore > 5), which could modulate NadR activity. However, since NadR inhibits *nadA* expression, an upregulation could only be achieved if nicotinate also inhibits NadR activity.

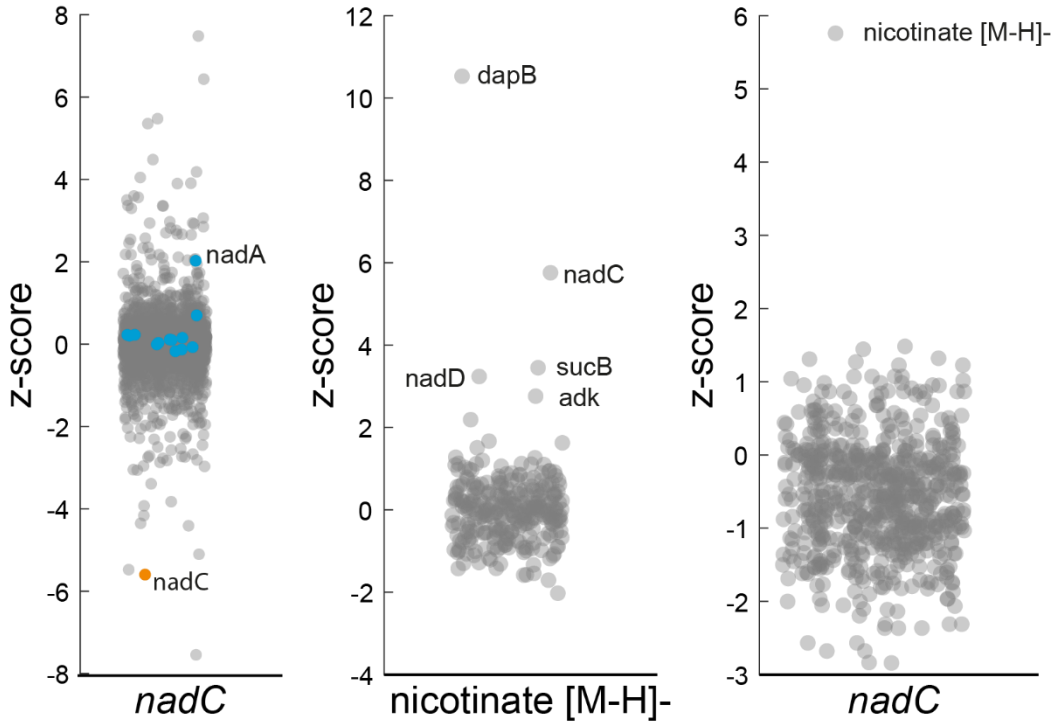


Figure 44. nadC and nicotinate. *(Left) proteome data of the nadC knockdown. Grey dots are all measured proteins. Blue dots are proteins of the corresponding kegg pathway. The orange dot is the target gene nadC. (Middle) Metabolome data. Dots are the 283 CRISPRi strains. (Right). Metabolome data. Dots are annotated metabolites.*

In conclusion, untargeted metabolomics showed that the CRISPRi knockdowns offset the metabolome. We identified substrate accumulations for 66 target genes. We showed that we can recover known metabolite regulators for the transcription factors ArgR and MetR. Moreover, we identified a putative interaction between the metabolite nicotinate and the transcription factor NadR that could control the expression of the gene *nadA* in the nicotinate and nicotine amid KEGG pathway. Next, we will identify further putative

metabolite-TF interactions in this dataset and then verify the candidates by orthogonal approaches like *in vitro* binding assays.

6.8 Discussion

Here, we measured the proteome and the metabolome of 283 CRISPRi knockdowns. In total, 45 % of the target genes had a z-score < -3, and 74 % of the target genes had a z-score <-1, indicating functional CRISPRi knockdowns. Besides the respective target genes, on average 40 other proteins showed strong changes. We hypothesized that some of these changes happen due to transcriptional feedback. Using a pathway-based approach we identified feedback regulation in 23 KEGG pathways. Our analysis showed that the genes *menH* and *folM* were upregulated although there was no known mechanism to cause an upregulation, indicating a knowledge gap in the TRN. In the literature we found 134 known metabolite-TF interactions and we used a distance analysis to show that effector metabolites are usually in proximity of the regulated genes. Untargeted metabolomics showed that the CRISPRi knockdowns offset the metabolome, as indicated by substrate accumulations of 66 CRISPRi targets. We showed for the transcription factors ArgR and MetR that it is possible to recover known metabolite regulators from the metabolomics data. Finally, the data suggested nicotinate as a candidate metabolite to control activity of the TF NadR. Thus, using a CRISPRi library to perturb 283 target genes in E. coli metabolism allowed us to map feedback regulation between metabolism and transcription.

In this study we integrated proteome data in a pathway-based approach to map feedback regulation in *E. coli*. This approach allowed us to find feedback regulation in 23 CRISPRi strains. Moreover, this approach indicated knowledge gaps in the TRN of E. coli, as indicated for the genes *menH* and *folM*. These results highlight the modular organisation of *E. coli* metabolism. Thus, a pathway-based approach enables the simultaneous identification of transcriptional feedback and knowledge gaps in the TRN. In the next steps, it will be important to assess the specificity of the proteome changes to ensure that changes are based on local responses and not global regulation.

To investigate the effects of the CRISPRi knockdowns on the metabolome, we measured the metabolome using FI-MS. We showed that on average twelve metabolites changed per CRISPRi strain and that for 66 CRISPRi strains the substrates accumulated. To distinguish between substrates and products we used FBA. We then showed for the transcription factors ArgR and MetR that it is possible to identify their metabolite effectors. Finally, we identified a putative feedback loop in the nicotinate and nicotine amid KEGG pathway. The knockdown of *nadC* caused accumulation of *nadA* that is regulated by NadR. NadR, however, is only known to be regulated by ATP and inspection of the metabolome data showed a specific nicotinate accumulation. To cause an upregulation of *nadA*, nicotinate needs to inhibit NadR activity. Next, it will be important to verify this interaction by orthogonal approaches, like *in vitro* binding assays.

In conclusion, we mapped transcriptional feedback regulation in *E. coli* using a metabolism wide CRISPRi library. We demonstrated that this approach allows us to identify knowledge gaps in the TRN, as well as to identify putative metabolites that control TF activity, and thereby gene expression *in vivo*.

6.9 References

(1) Stelling, J.; Sauer, U.; Szallasi, Z.; Doyle, F. J.; Doyle, J. Robustness of Cellular Functions. *Cell* **2004**, *118* (6), 675–685. https://doi.org/10.1016/j.cell.2004.09.008.

(2) Kitano, H. Biological Robustness. *Nat Rev Genet* **2004**, *5* (11), 826–837. https://doi.org/10.1038/nrg1471.

(3) Chubukov, V.; Gerosa, L.; Kochanowski, K.; Sauer, U. Coordination of Microbial Metabolism. *Nat Rev Microbiol* **2014**, *12* (5), 327–340. https://doi.org/10.1038/nrmicro3238.

(4) Kochanowski, K.; Sauer, U.; Chubukov, V. Somewhat in Control--the Role of Transcription in Regulating Microbial Metabolic Fluxes. *Curr Opin Biotechnol* **2013**, *24* (6), 987–993. https://doi.org/10.1016/j.copbio.2013.03.014.

(5) Sander, T.; Farke, N.; Diehl, C.; Kuntz, M.; Glatter, T.; Link, H. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. Coli by Enforcing Enzyme Overabundance. *Cell Syst* **2019**, *8* (1), 66-75.e8. https://doi.org/10.1016/j.cels.2018.12.005.

(6) Reznik, E.; Christodoulou, D.; Goldford, J. E.; Briars, E.; Sauer, U.; Segrè, D.; Noor, E. Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Rep* **2017**, *20* (11), 2666–2677. https://doi.org/10.1016/j.celrep.2017.08.066.

(7) Piazza, I.; Kochanowski, K.; Cappelletti, V.; Fuhrer, T.; Noor, E.; Sauer, U.; Picotti, P. A Map of Protein-Metabolite Interactions Reveals Principles of Chemical

Communication. *Cell* **2018**, *172* (1–2), 358-372.e23. https://doi.org/10.1016/j.cell.2017.12.006.

(8) Umbarger, H. E. Evidence for a Negative-Feedback Mechanism in the Biosynthesis of Isoleucine. *Science* **1956**, *123* (3202), 848–848. https://doi.org/10.1126/science.123.3202.848.a.

(9) Dieckmann, R.; Pavela-Vrancic, M.; von Döhren, H.; Kleinkauf, H. Probing the Domain Structure and Ligand-Induced Conformational Changes by Limited Proteolysis of Tyrocidine Synthetase 1. *J Mol Biol* **1999**, *288* (1), 129–140. https://doi.org/10.1006/jmbi.1999.2671.

(10) Diether, M.; Sauer, U. Towards Detecting Regulatory Protein–Metabolite Interactions. *Current Opinion in Microbiology* **2017**, *39*, 16–23. https://doi.org/10.1016/j.mib.2017.07.006.

(11) Reinhard, F. B. M.; Eberhard, D.; Werner, T.; Franken, H.; Childs, D.; Doce, C.; Savitski, M. F.; Huber, W.; Bantscheff, M.; Savitski, M. M.; Drewes, G. Thermal Proteome Profiling Monitors Ligand Interactions with Cellular Membrane Proteins. *Nat Methods* **2015**, *12* (12), 1129–1131. https://doi.org/10.1038/nmeth.3652.

(12) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Martinez Molina, D.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science* **2014**, *346* (6205), 1255784. https://doi.org/10.1126/science.1255784.

(13) Nikolaev, Y. V.; Kochanowski, K.; Link, H.; Sauer, U.; Allain, F. H.-T. Systematic Identification of Protein-Metabolite Interactions in Complex Metabolite Mixtures by Ligand-Detected Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2016**, *55* (18), 2590–2600. https://doi.org/10.1021/acs.biochem.5b01291.

(14) Hackett, S. R.; Zanotelli, V. R. T.; Xu, W.; Goya, J.; Park, J. O.; Perlman, D. H.; Gibney, P. A.; Botstein, D.; Storey, J. D.; Rabinowitz, J. D. Systems-Level Analysis of Mechanisms Regulating Yeast Metabolic Flux. *Science* **2016**, *354* (6311), aaf2786. https://doi.org/10.1126/science.aaf2786.

(15) Gerosa, L.; Haverkorn van Rijsewijk, B. R. B.; Christodoulou, D.; Kochanowski, K.; Schmidt, T. S. B.; Noor, E.; Sauer, U. Pseudo-Transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Syst* **2015**, *1* (4), 270–282. https://doi.org/10.1016/j.cels.2015.09.008.

(16) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.

(17) Anglada-Girotto, M.; Handschin, G.; Ortmayr, K.; Campos, A. I.; Gillet, L.; Manfredi, P.; Mulholland, C. V.; Berney, M.; Jenal, U.; Picotti, P.; Zampieri, M. Combining CRISPRi and Metabolomics for Functional Annotation of Compound Libraries. *Nat Chem Biol* **2022**, *18* (5), 482–491. https://doi.org/10.1038/s41589-022-00970-3.

(18) Donati, S.; Kuntz, M.; Pahl, V.; Farke, N.; Beuter, D.; Glatter, T.; Gomes-Filho, J. V.; Randau, L.; Wang, C.-Y.; Link, H. Multi-Omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms That Buffer Decreases of Enzymes in E. Coli Metabolism. *Cell Systems* **2021**, *12* (1), 56-67.e6. https://doi.org/10.1016/j.cels.2020.10.011.

(19) Scott, M.; Gunderson, C. W.; Mateescu, E. M.; Zhang, Z.; Hwa, T. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **2010**, *330* (6007), 1099–1102. https://doi.org/10.1126/science.1192588.
(20) Scott, M.; Hwa, T. Bacterial Growth Laws and Their Applications. *Curr Opin Biotechnol* **2011**, *22* (4), 559–565. https://doi.org/10.1016/j.copbio.2011.04.014.
(21) Scott, M.; Klumpp, S.; Mateescu, E. M.; Hwa, T. Emergence of Robust Growth Laws from Optimal Regulation of Ribosome Synthesis. *Mol Syst Biol* **2014**, *10*, 747. https://doi.org/10.15252/msb.20145379.
(22) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **2000**, *28* (1), 27–30. https://doi.org/10.1093/nar/28.1.27.
(23) Lempp, M.; Farke, N.; Kuntz, M.; Freibert, S. A.; Lill, R.; Link, H. Systematic Identification of Metabolites Controlling Gene Expression in E. Coli. *Nat Commun* **2019**, *10* (1), 4463. https://doi.org/10.1038/s41467-019-12474-1.
(24) Salgado, H.; Gama-Castro, S.; Peralta-Gil, M.; Díaz-Peredo, E.; Sánchez-Solano, F.; Santos-Zavaleta, A.; Martínez-Flores, I.; Jiménez-Jacinto, V.; Bonavides-Martínez, C.; Segura-Salazar, J.; Martínez-Antonio, A.; Collado-Vides, J. RegulonDB (Version 5.0): Escherichia Coli K-12 Transcriptional Regulatory Network, Operon Organization, and Growth Conditions. *Nucleic Acids Res* **2006**, *34* (Database issue), D394-397. https://doi.org/10.1093/nar/gkj156.
(25) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat Biotechnol* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.
(26) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; Feist, A. M.; Palsson, B. O. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat Biotechnol* **2017**, *35* (10), 904–908. https://doi.org/10.1038/nbt.3956.
(27) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat Biotechnol* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.

Chapter 7 A network approach identifies in-source modifications of primary metabolites during flow-injection mass spectrometry

Niklas Farke[#], Thorben Schramm[#], Andreas Verhülsdonk, Hannes Link.

[#]Authors contributed equally

This chapter is written in manuscript style and is currently under revision at ACS Analytical Chemistry.

I performed experiments and data analysis, developed the analysis software, created the figures, wrote the manuscript and designed the study.

7.1 Abstract

Flow-injection mass spectrometry (FI-MS) enables metabolomics studies with a very high sample-throughput. In most FI-MS methods, samples are directly injected into the electrospray ionization (ESI) source of a high-resolution mass spectrometer, and metabolites are annotated to ion peaks in the MS1 spectrum based on their exact mass over charge. Although ESI is considered a soft ionization technique, it can cause in-source modifications of analytes that are then misannotated to metabolites. In-source modifications include adduct formation, fragmentation, and other chemical reactions of metabolites. These effects are especially prominent in FI-MS because all analytes and the sample matrix enter the ESI at the same time. Here, we spiked authentic standards of 160 primary metabolites individually into an Escherichia coli metabolite extract and measured the thus derived 160 spike-in samples by FI-MS. Out of the 160 metabolites, 154 were annotated in their protonated or deprotonated form to ion peaks in the MS1 spectrum, and 134 of these ion peaks increased in the respective spike-in the standard. These results demonstrated that FI-MS can capture a wide-range of chemically diverse analytes within 30 seconds measurement time. However, the data also revealed extensive in-source modifications: across all 160 spike-in samples, we identified significant increases of 11,013 ion peaks in positive and negative mode combined. To explain these unknown *m/z* features, we connected them to the *m/z* feature of the (de-)protonated metabolite using information about mass differences and MS2 spectra. This resulted in networks that explained on average 49% of all significant *m/z* features. The networks showed that a single metabolite undergoes compound specific and often sequential in-source modifications like adductions, chemical reactions, and fragmentations. Taken together, our results show that FI-MS generates complex MS1 spectra, which can lead to a 68-fold overestimation of significant features. Yet, known mass differences and MS2 level information can explain these features and can therefore avoid misannotation of metabolites in FI-MS analyses

7.2 Introduction

Flow-injection mass spectrometry (FI-MS) does not rely on chromatographic separation of analytes[1,2]. Instead, samples are injected into the mobile phase that directly enters a mass spectrometer. Metabolites are then distinguished solely by their mass to charge ratio (*m/z*) in the MS1 spectrum. This makes FI-MS faster than methods with chromatographic separation[3,4] and enables run times on the second-time scale or even real-time metabolomics with living cells[5].

FI-MS has been applied to measure the metabolome in various organisms including *Escherichia coli*, yeast, ruminants, and human cancer cell lines[6–10]. In these studies, hundreds or even thousands of strains or conditions could be analyzed due to the fast measurement time of FI-MS. Although FI-MS detects usually a very large number of *m/z* features (ion peaks in the MS1 spectrum), only a small fraction of *m/z* features can be annotated to metabolites. Thus, there is a large number of unexplained *m/z* features in FI-MS analyses, which could mean that either many metabolites are not known or that single metabolites produce multiple *m/z* features. Annotation of unknown *m/z* features is a general challenge in all untargeted metabolomics methods[11–14]. For example, an untargeted LC-MS analysis suggested that out of 25,000 measured *m/z* features less than 1,000 originated from unique metabolites[15]. The high number of *m/z* features in untargeted metabolomic methods is often attributed to contaminants, isotopes, modification of metabolites in the ion-source, and other mass spectrometry artifacts.

In-source fragmentation is one example of such mass spectrometry artifacts that increase the number of *m/z* features per metabolite. The conditions in the ESI can lead to fragmentation because metabolites are subjected to high temperatures (150°C to 400°C) and electric potentials between 2000 V and 4000 V. While ESI sources are usually designed to minimize in-source fragmentation, it is also possible to promote in-source fragmentation such that MS1 spectra resemble MS2 spectra that were obtained by collision induced dissociation[16]. Apart from in-source fragmentation, other modifications of metabolites in the ion-source includes the formation of adducts (e.g. with Na, K, ammonia, sulfate), gains or losses of functional groups by chemical reactions

(methylation, phosphorylation), or formation of homo- and heterodimers. Even self-cyclization has been observed for glutamate and glutamine[17].

A common approach to identify in-source modifications and improve *m/z* feature annotation is based on chromatographic peak shape correlation analysis[18–20]. This approach considers that *m/z* features from the same metabolite must have the same elution profile[21]. Chromatographic peak shape correlation analysis is especially effective if it is combined with MS2 spectra[19,22,23] or isotope labeled substrates[11,21]. Some recent molecular networking approaches[19,23,24] combine similarities of elution profiles and MS2 spectra to identify in-source modifications and to increase annotation confidence. In isotope labelling approaches, metabolites are labelled by feeding cells with ^{13}C-carbon or ^{15}N-nitrogen sources[11,21], which changes the mass of all metabolites (N- or C-containing) but not their retention times. Analyzing the mass differences of *m/z* features with the same retention time can then improve annotation confidence and identification of in-source modifications or contaminants.

Because FI-MS lacks a chromatographic separation, it is not possible to detect in-source effects by chromatographic peak shape correlation analysis. Therefore, approaches to consider in-source effects in FI-MS are limited and currently based on extending the list of reference masses[1,25]. This means that, instead of annotating *m/z* features only to (de-)protonated metabolites, they are also annotated to the most prevalent adducts and neutral losses or gains. However, this approach cannot identify complex sequential in-source modifications due to combinatorial explosion of the reference list. Moreover, it is difficult to unequivocally annotate *m/z* features to a single entry in a reference list, especially if they include a large number of metabolites and derivatives with the same mass.

Here, we used an experimental approach to identify in-source modifications of metabolites in FI-MS. Therefore, we spiked 160 metabolite standards individually into an *E. coli* extract and measured MS1 spectra by FI-MS. We then searched for *m/z* features that increased in a spike-in sample relative to all other spike-in samples. On average 68 *m/z* features increased per spike-in standard suggesting extensive in-source modifications. While some spike-in standards showed hundreds of significant *m/z* features that should all originate from a single metabolite standard, others showed only

increases of the *m/z* feature that matched the (de-)protonated metabolite standard. We could explain 49% of the significant *m/z* features by connecting them in networks that represent known in-source reactions, adducts, isotope patterns, and in-source fragments.

7.3 FI-MS with 160 authentic metabolite standards

We prepared 160 authentic standards of primary metabolites and spiked them individually into a metabolite extract from glucose-fed *E. coli* cells (**Supporting Information: Table A**). The 160 standards fall into six functional categories: amino acid metabolism, nucleotide metabolism, central metabolism, cofactor metabolism, antioxidants, and others. Each metabolite standard was added to the *E. coli* metabolite extract at a final concentration of 1 µM and was measured by FI-MS in both positive and negative ionization mode (three analytical replicates) (**Figure *45*a**).

Out of 160 metabolite standards, 154 were annotated in their protonated or deprotonated form to an ion peak in the MS1 spectrum (**Figure *45*b**). Six metabolites were not annotated, either due to low abundant ion peaks (< 1,000 counts: menadione, 3,4-dihydroxy-L-phenylalanine, tetrahydrofolic acid, carbamoyl-P, and L-cysteine) or because the ion peak prominence was too low (< 1,000 counts: argininosuccinic acid). Next, we inspected if the addition of a metabolite standard led to increases of the respective ion peak. For example, spike-in samples with ATP, GTP, CTP, and UTP showed increases of ion peaks that matched the protonated and deprotonated form of these metabolites (**Figure *45*c**). Notably, increases of all nucleotides were consistent between three analytical replicates, showing that FI-MS is reproducible. Ion peaks of ATP, GTP, CTP, and UTP were also present in the other spike-in samples (black spectra in **Figure *45*c**), but the corresponding ion intensities were often low and close to the baseline signal. These "near-baseline" ion peaks of ATP, GTP, CTP, and UTP were not present in a ^{13}C-labeled *E. coli* extract, thus confirming that these peaks originate from endogenous *E. coli* nucleotides (**Appendix: Figure 54**). In 134 spike-in samples, the ion peaks that matched the (de-)protonated metabolites were significantly increased in either ionization mode (z-score > 3, **Figure *45*d, Supporting Information: Table D**). In

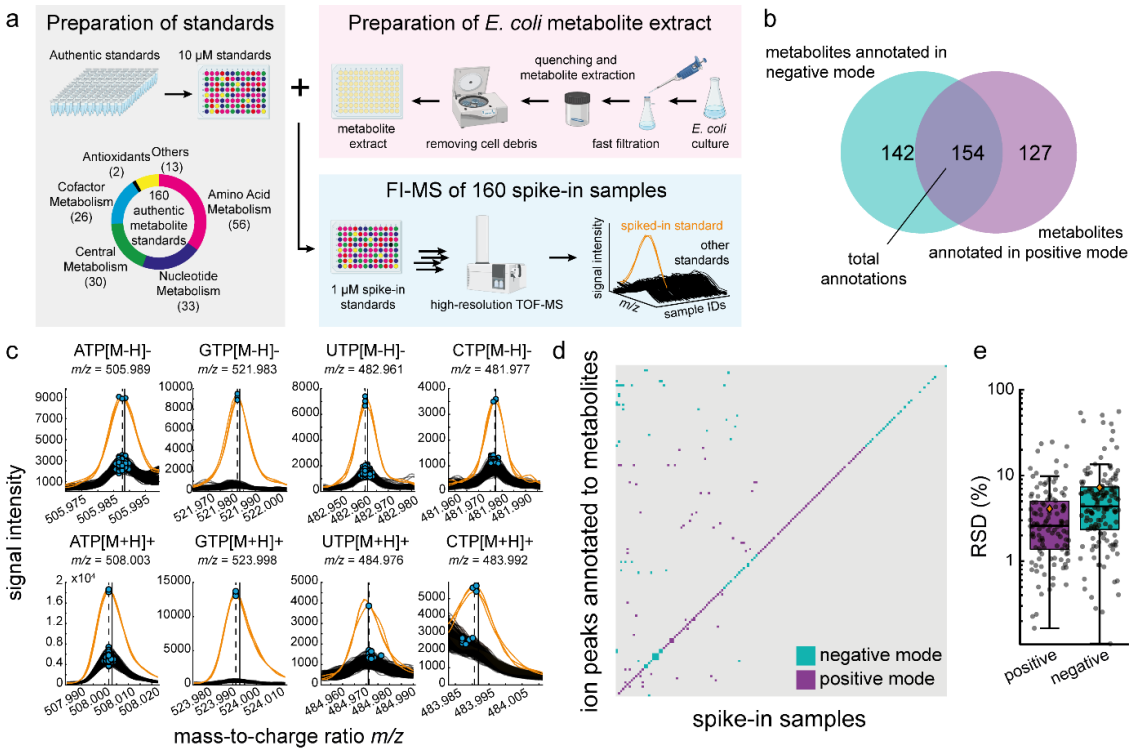


Figure 45. ***FI-MS with 160 authentic metabolite standards. a)*** *160 authentic metabolite standards were spiked into an E. coli metabolite extract at a final concentration of 1 µM. Spike-in samples were measured by FI-MS in analytical triplicates in positive and negative ionization mode. BioRender.com was used to create this figure* ***b)*** *Number of ion peaks (m/z features) that were annotated to the protonated (positive mode) and deprotonated (negative mode) form of the 160 metabolite standards.* ***c)*** *Ion peaks that are annotated to four nucleotides (ATP, CTP, GTP, UTP) in positive and negative ionization mode. The spike-in sample that contains the respective nucleotide is indicated in orange, the other 159 spike-in samples are black. Blue dots indicate m/z features in single samples, and vertical dotted lines are merged and centroided m/z features. Vertical solid lines indicate the monoisotopic masses of the nucleotides plus/minus the mass of a proton.* ***d)*** *The binary heatmap shows increases of m/z features that are annotated to metabolite standards in the spike-in samples. Significant increases of m/z features (z-score > 3) are shown in blue (negative mode) and purple (positive mode). Columns are the spike-in samples and rows the respective m/z features.* ***e)*** *Boxplots show the relative standard deviation (RSD) of metabolite standards. Black dots are the RSD for each spike-in metabolite (n = 3). Orange diamonds are the means.*

negative ionization mode, 120 spike-in samples showed increased signals as deprotonated metabolites ([Metabolite-H]-). In positive ionization mode, 105 peaks increased as protonated metabolites ([Metabolite+H]+). In the following, we will refer to significantly changed ion peaks with a z-score >3 as “significant features”. 26 spike-in standards did not show a significant feature at the (de-)protonated ion peak. One explanation for this is that the metabolites have already a high concentration in the *E. coli* metabolite extract and that an addition of 1 µM does not lead to a strong increase with a z-score > 3. For example, reduced glutathione is one of the most abundant

metabolites in *E. coli*[28], and the addition of glutathione standard hardly increased its concentration in the spike-in sample (**Appendix: Figure 55**).

FI-MS was reproducible because the median relative standard deviation (RSD) between the three analytical replicates was below 5 % for signals from metabolite standards in negative and positive ionization mode (**Figure *45*e**). Signals from endogenous metabolites, had a median RSD of 12.4 % in positive and 19.4 % negative mode (**Appendix: Figure 56**).

These results suggested that FI-MS can detect concentration changes of chemically diverse metabolites, which are in the physiological range of intracellular metabolites (1 µM in the final sample corresponds to ca. 1 mM intracellularly). However, we noticed that many significant features did not match the (de-)protonated form of the metabolite in the spike-in sample (e.g. significant features that are off the diagonal in **Figure *45*d**). Thus, we next investigated all significant features in all spike-in samples.

7.4 Single metabolites can produce extensive in-source derivates

Most spike-in samples showed significant features (ion peaks with a z-score > 3) that matched the protonated or deprotonated metabolite (**Figure *45*d**). However, most spike-in samples had more significant features than only the (de-)protonated metabolite standard (**Figure 46a**). On average, we found 68 significant features per spike-in sample, and 11 spike-in samples showed more than 100 significant features. The glycerol 3-phosphate(G3P) spike-in sample had the highest number of significant features

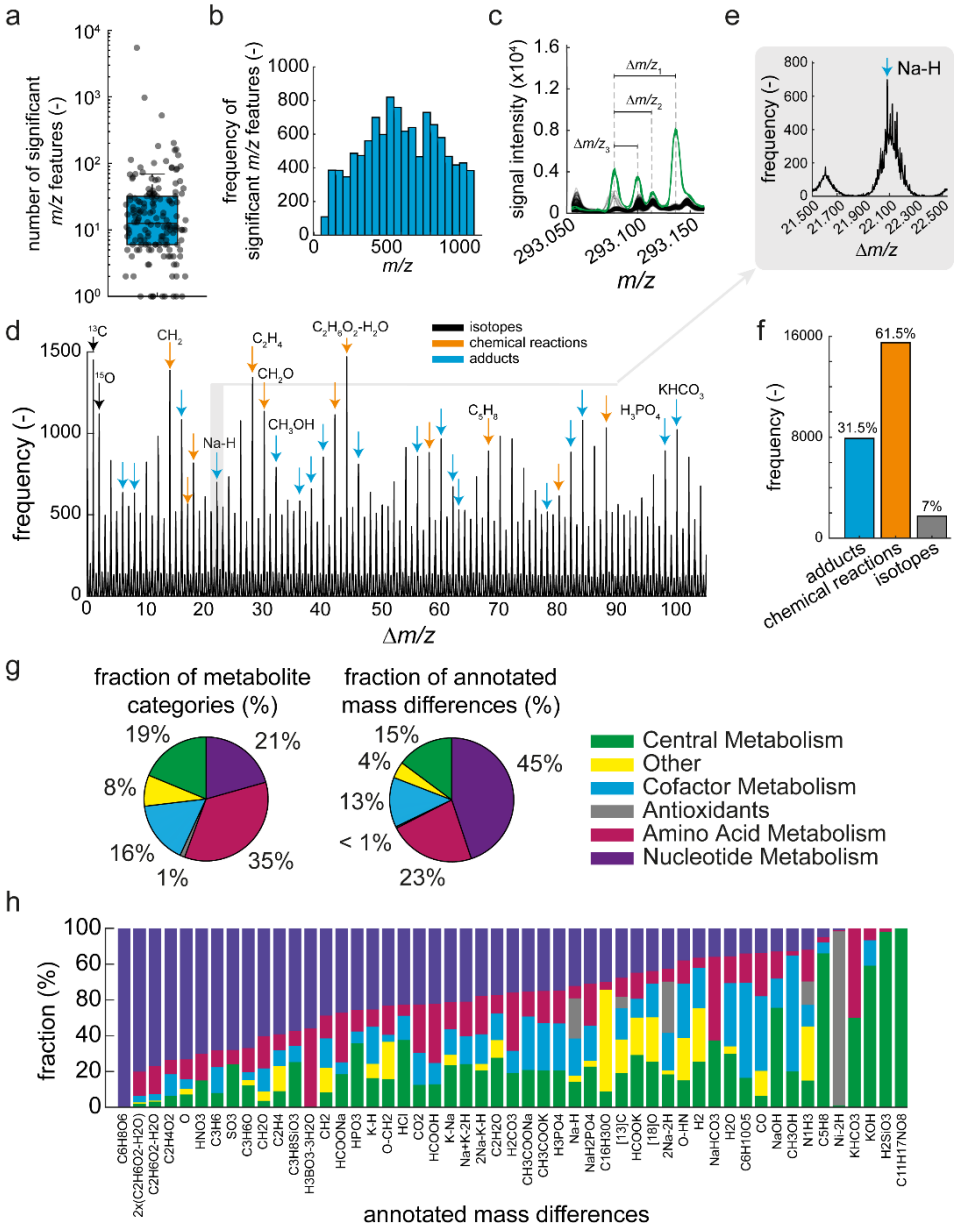


Figure 46. Systematic analysis of all m/z features that increase in spike-in samples. a) *Number of m/z features with a z-score > 3 (significant m/z features) in each of the 160 spike-in samples (grey dots). The upper and lower edges of the box in the boxplot indicate the 25 % and 75 % percentiles, and the line is the median.* ***b)*** *Histogram showing the distribution of significant m/z features over the MS1 spectrum.* ***c)*** *Example of significant m/z features in the MS1 spectrum (100 mDa window) of the spike-in sample with glycerol 3-phosphate. Green lines are the glycerol 3-phosphate spike-in samples (n = 3), and black lines are the other spike-in samples. Rulers indicate the mass differences between two m/z features in the spectra.* ***d)*** *Δm/z spectrum based on the pairwise mass differences between all significant m/z features in all 160 spike-in samples (shown is the Δm/z range between 0 Da and 110 Da). The peak height corresponds to the frequency of a Δm/z value. Arrows indicate Δm/z peaks that match mass differences of known isotopes, chemical reactions or adducts.* ***e)*** *Example of the Δm/z peak that matches the sodium adduct [Na-H].* ***f)*** *Fraction of Δm/z peaks that match known isotopes, chemical reactions or adducts.* ***g)*** *The left pie chart shows the fraction of metabolite categories across the 160 standards. The right pie chart shows the fraction of annotated mass differences for each metabolite category.* ***h)*** *Stacked bar plot showing the relationship between the functional categories of the spiked-in metabolites and the annotated mass differences. The fraction indicates the ratio between the number of spike-in metabolites of a specific category, in which the mass difference occurred, and the total number of samples, in which the mass difference occured. The spike-in samples of glycerol 3-phosphate and fumarate were left out.*

(5,464, **Figure 46a**). Across all 160 spike-in samples, FI-MS in positive mode showed 10,206 significant features and 807 in negative mode. The significant features were distributed over the entire mass spectrum and occurred even in the higher mass range of 800 – 1,000 *m/z* (**Figure 46b**). To understand the origin of these significant features, we first calculated the mass differences (Δm/z) between all pairs of significant features in a single spike-in sample (**Figure 46c**). Several mass differences (Δm/z) occurred frequently across the 160 spike-in samples, thus indicating common in-source effects like neutral losses, adduct formation, and chemical reactions that are prevalent for many different compounds (**Figure 46d and e**). 51 mass differences that appeared more than ten times matched known in-source effects and isotope pattern reported in the literature23. Out of these 51 known mass differences, 23 were chemical reactions, 26 were adducts, and 2 were natural isotopes (13C and 18O). The 23 chemical reactions account for 61.5 % of the frequent mass differences (>10 times in all samples), the 22 adducts for 31.5%, and the isotopomers containing 13C or 18O for 7 % (**Figure 46f**). For example, the 21.982 Da mass difference of a Na-H neutral loss occurred in total 699 times and was among the most frequent ones (**Figure 46e**). The ten most frequent mass differences occurred more than 1,000 times across all 160 spike-in samples, and eight of them could be explained with the mass differences in the literature (**Figure 46d**). We then wondered whether certain mass differences occurred more frequently for metabolites of a specific functional category than for metabolites of other categories. For example, only 21 % of the 160 standards were metabolites from nucleotide metabolism. Yet, they accounted for 45 % of all explainable mass differences (**Figure 46g**). This indicated that metabolites from nucleotide metabolism were more susceptible to modifications than metabolites in other categories. In contrast, 35 % of the 160 metabolite standards were part of amino acid metabolism but they covered only 23 % of the explainable mass differences indicating that metabolites from amino acid biosynthesis were less prone to modifications in our reference list of mass differences than the other metabolites. Metabolites from central metabolism as well as cofactor biosynthesis accounted for 19 % and 16 % of the standard library, respectively, and they explained a similar fraction of mass differences (15 % and 13 %).

Since some metabolite categories were more often modified than others, we looked into individual mass differences and examined whether certain mass differences occurred preferably for specific categories (**Figure 46h**). Indeed, the data indicated that individual mass differences were more frequent for some categories than for others. For example, the $C_6H_8O_6$ neutral loss occurred exclusively in metabolites from nucleotide metabolism. Similarly, many other modifications, including modifications with O, HNO_3, $C_2H_4O_2$, or SO_3, occurred more frequently with metabolites from nucleotide metabolism.

7.5 A network approach explains significant *m/z* features in FI-MS spectra

We expected that the significant features can be linked to the (de-)protonated spike-in metabolite by single and multiple modification steps. To test this, we created a network for each spike-in sample, in which nodes represent all significant features. Then, we drew an edge between two nodes if the mass difference $\Delta m/z$ between them matched one of the 51 frequent mass differences (the mass differences identified above, see **Figure 46h**). Thus, edges represent in-source effects and nodes significant features (see schematic in **Figure 47a**).

The thus derived networks connected on average 43 % of the significant features in a spike-in sample to the *m/z* feature of the respective (de-)protonated metabolite (**Figure 47a**). Thus, 43 % of the significant features can be linked to a single metabolite and therefore, are explained by the 51 frequent mass differences. Only 20 % of the significant features were directly linked to the *m/z* feature of the (de-)protonated metabolite (**Figure 47a**). This shows that single in-source modifications account only for half of the significant features and that sequential modifications are frequent.

For example, the glucosamine 6-phosphate (Ga6p) spike-in standard showed five significant features, which are all directly or indirectly connected to the *m/z* feature that matches protonated Ga6p (**Figure 47b**). Two significant features are directly connected

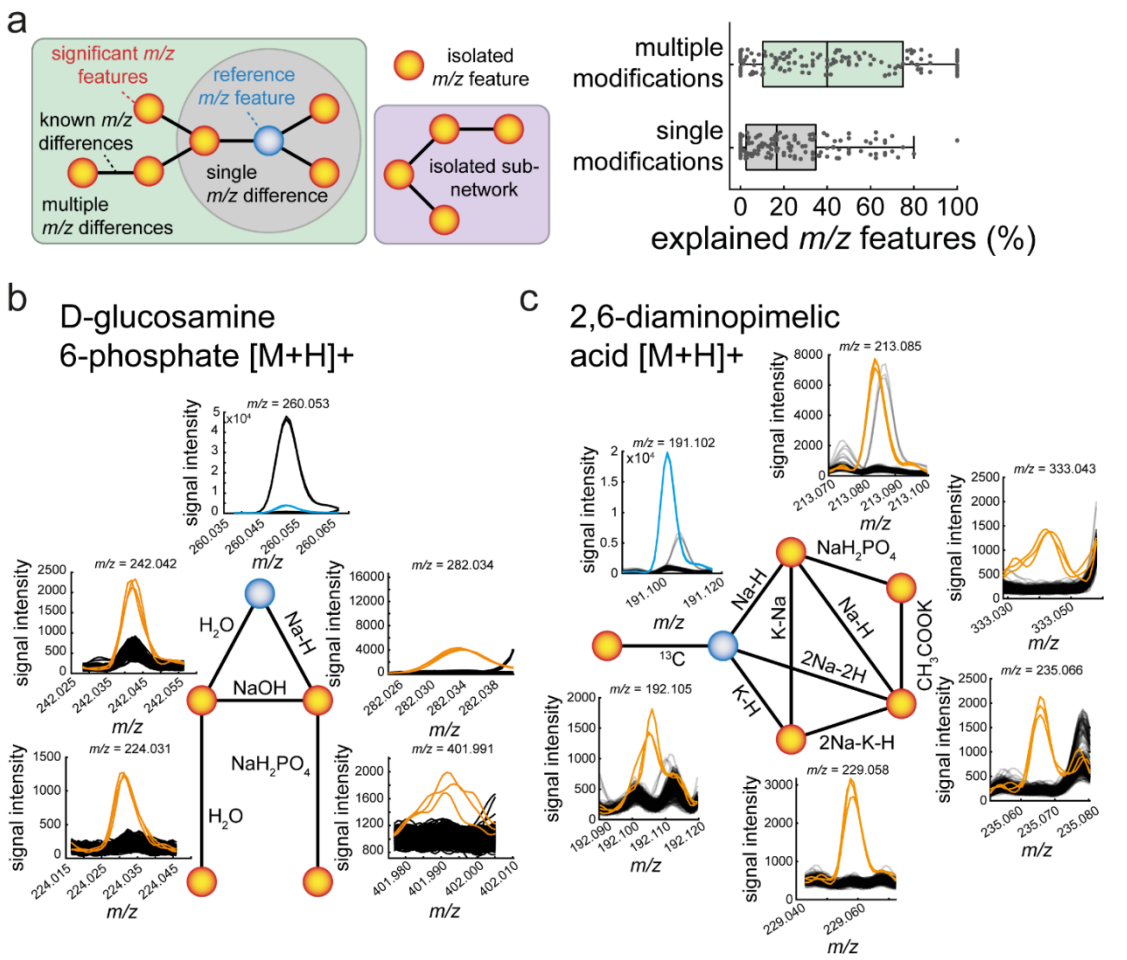


Figure 47. Networks of significant m/z features. a) *Concept figure showing the structure of a m/z feature network: orange nodes are significant m/z features. The blue node is a significant m/z feature that is annotated to the (de-)protonated metabolite standard (reference m/z feature). Nodes are connected by edges that correspond to one of 51 known m/z differences. The grey circle indicates nodes that are directly connected to the reference m/z feature by a single m/z difference. The green box indicates a network, in which all nodes are connected with the reference node including multiple sequential combinations of m/z differences. Isolated sub-networks are not connected to the reference node (purple box). Isolated m/z features are not connected to any other node. The boxplots show the fraction of significant m/z features that are connected to the reference m/z feature by a single m/z difference or by multiple sequential combinations of m/z differences.* ***b)*** *Example of the feature network of the spike-in sample with D-glucosamine 6-phosphate (Ga6p).* ***c)*** *Same as b) for the diaminopimelic acid (DAP) spike-in sample in positive ionization mode.*

to the protonated Ga6p mass, and they are likely a water loss (H_2O) and a sodium adduct (Na-H). Two other significant features (*m/z* = 224.031 and *m/z* = 242.042) were two steps away from protonated Ga6P, and they were explained by a double loss of water and a NaH_2PO_4 adduction to the sodium adduct. Thus, drawing edges in an unbiased way between all pairs of nodes resulted in a network that explained all significant features of the Ga6p spike-in sample.

In many networks, the nodes (significant *m/z* features) were connected to the (de-)protonated spike-in metabolite by different series of sequential modifications. One specific series of sequential modifications could be an initial modification by Na-K that is followed by a second modification like Na-H. In some cases, different series of sequential modifications have very similar net mass changes and can explain the same significant feature. One example that illustrates this phenomenon is the feature network of 2,6-diaminopimelic acid (DAP) in positive ionization mode. The DAP spike-in sample showed six significant features, which were all connected with the *m/z* feature of protonated DAP (**Figure 47c**).

The network approach connected on average 43 % of all significant features of a spike-in sample to the (de-)protonated metabolite standard. Yet, some significant features had no connection to others or they formed sub-networks with no connection to the (de-)protonated metabolite (see schematic in **Figure 47a**). Therefore, we next examined whether sub-networks and isolated features are caused by in-source fragmentation, which can lead to similar effects as collision induced dissociation in tandem mass spectrometry[16].

7.6 MS2 information identifies significant features that are in-source fragments

To identify significant features that originate from in-source fragmentation of the metabolite standard, we used information about MS2 spectra in the human metabolome database (HMDB[27]). HMDB listed experimental MS2 spectra for 152 out of 160 metabolite standards (**Figure 48a**). 103 standards had at least one significant feature that matched an ion peak in the MS2 spectrum). On average, each spike-in sample had 3.4 in-source fragments indicating a substantial number of in-source fragmentation events during FI-MS (**Figure 48a**). In total, 551 MS2 features matched the significant *m/z* features. As expected, in-source fragments had masses in the lower *m/z* range, between 50 and 500 *m/z* (**Figure 48b**). One example of an in-source fragment is hypoxanthine, which is formed by fragmenting inosine and inosine monophosphate (IMP). Consequently, the ion peak of hypoxanthine increased in the IMP and inosine spike-in

samples (**Figure 48c**). Then, we tested whether fragments were present in sub-networks without a link to the metabolite standard. For example, the L-citrulline spike-in sample showed 8 significant

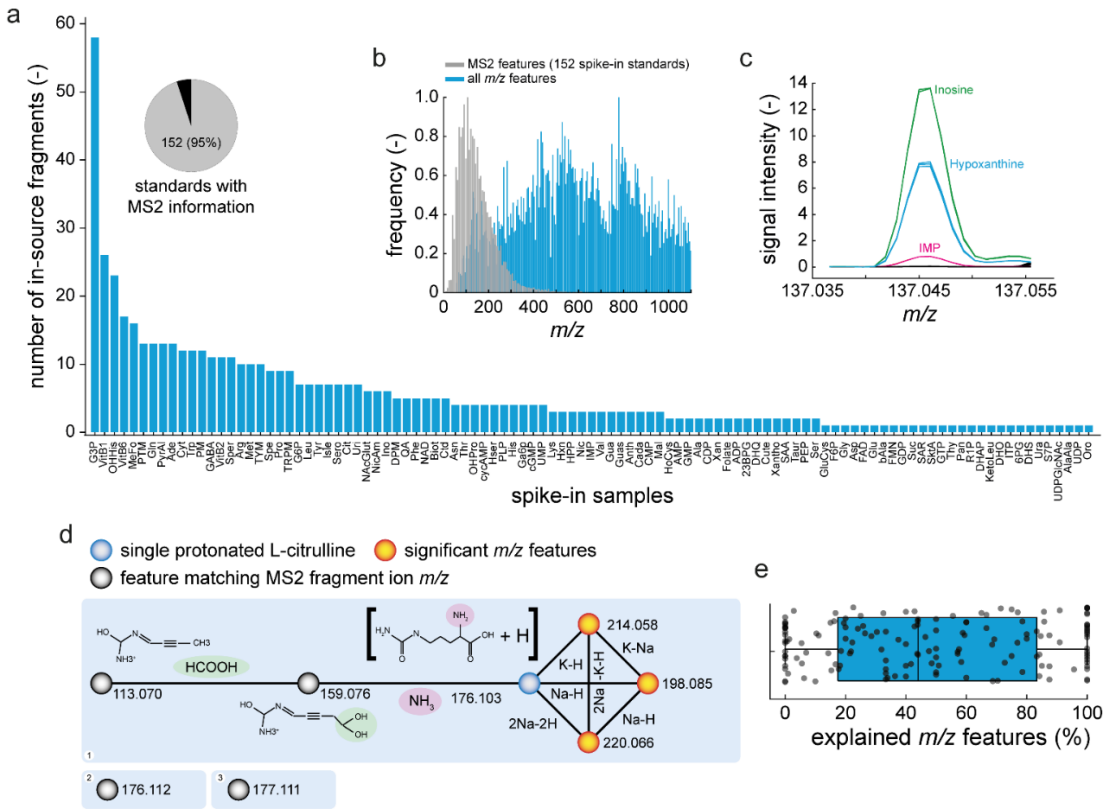


Figure 48. Identification of in-source fragments by MS2 spectra of metabolite standards. a) *The pie chart shows the fraction of metabolite standards with MS2 information in the human metabolome database (HMDB). The bar plot shows the number of MS2 fragment masses that match the features of individual spike-in samples.* ***b)*** *Histogram showing the distribution of the m/z features. Blue are all significant m/z features from the spike-in samples. Grey are MS2 fragment masses.* ***c)*** *Example MS1 spectrum at the mass of hypoxanthine. The purple line is the IMP spike-in sample, the blue line is the hypoxanthine spike-in sample, and the green line is the inosine spike-in sample.* ***d)*** *Example network of L-citrulline. Nodes are features and edges are explained mass differences. Black nodes are features that matched MS2 fragment masses from HMDB. The blue node is the protonated mass of citrulline. The orange nodes are other features. For the features at m/z = 113.070 and m/z = 159.076, structures were predicted by CFM-ID.* ***e)*** *Boxplot showing the fraction of explained features for all spike-in samples. Each black point corresponds to a spike-in sample and shows the explained fraction of features. Upper and lower box edges indicate the 25 % and 75 % percentiles. The whiskers indicate the furthest point, at which samples were not considered as outliers. The black line indicates the median.*

features, 6 of which were connected with the *m/z* feature of the (de-)protonated metabolite (**Figure 48d**). Two significant features were isolated nodes with no connection, but they were in the MS2 spectrum of citrulline (*m/z* = 177.111 and *m/z* = 176.112). The MS2 spectrum included another two *m/z* features that were already linked to the metabolite standard (*m/z* = 113.070 and *m/z* = 159.076), thus indicating

that collision induced dissociation (CID) produces some of the 51 in-source modifications. In case of L-citrulline, these were a neutral loss of NH_3 and a neutral loss of a HCOOH group. To confirm that these losses occur by CID, we predicted fragment structures of L-citrulline by CFM-ID[29]. Indeed, CFM-ID predicted the fragment structures that matched the masses of the two *m/z* features (113.070 and 159.076) and confirmed the neutral losses of NH_3 and HCOOH (**Figure 48d**). In total, adding in source fragments to our networks explained another 6 % of significant m/z features. Thus, on average, 49 % of all significantly changed m/z features were connected with the correct metabolite standard either by known in-source modifications or by in-source fragmentation (**Figure 48e**).

7.7 Misannotation of in-source derivates to metabolites

A single metabolite can produce multiple significant features, and we wondered how many of these significant features were misannotated to a metabolite that was not spiked into the sample. To determine how many significant features were misannotated, we used a reference list of 961 *E. coli* metabolites from the genome-scale metabolic model iML1515[30]. Since FI-MS cannot resolve isomers, they were considered as a single metabolite.

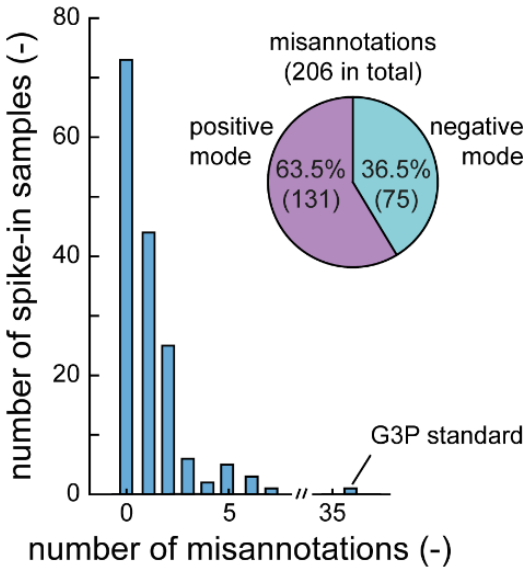


Figure 49. Misannotation of significant features to metabolites*. The histogram shows the number of significant features that are misannotated to a metabolite. The pie chart shows the fraction of misannotations for positive and negative ionization mode.*

In 54 % of our standards (87/160), at least one significant feature was falsely annotated to a metabolite (**Figure 49**). 18 standards had more than two misannotations. Overall, 64 % of the misannotations were in positive ionization mode (131 in total) and 37 % were in negative ionization mode (75 in total) (**Figure 49**). This means that biological screens with FI-MS are prone to misannotations if only (de-)protonated masses are considered. Based on our results, an estimate is that one (true) increase of a single metabolite will cause one (false) increase of an ion peak that is misannotated to another metabolite.

7.8 Discussion

FI-MS methods have been used for metabolome analyses in various studies[1,2,4–10]. Their advantages are fast analysis times (10 to 30 seconds per sample) and a high coverage of metabolites (often more than 1,000 putatively annotated metabolites[6,7]). Disadvantages, however, are low confidence levels of metabolite annotation and a high susceptibility to matrix effects due to the lack of chromatographic separation. Here, we confirmed the broad metabolite coverage of FI-MS, which detected increases of 134 out of 160 metabolite standards based on (de-)protonated ion peaks in the MS1 spectrum. However, we also observed pervasive in-source modifications of metabolites. These in-source modifications lead to multiple ion peaks per metabolite in the MS1 spectrum and, in the worst case, to false positive hits in FI-MS analyses of biological samples. By systematically analyzing FI-MS data from 160 spike-in standards, we found that a single metabolite produces on average 68 significant features and that, in extreme cases, more than 1,000 significant features originate from only one metabolite. This observation matches previous LC-MS based studies where the majority of *m/z* features were attributed to in-source modifications and only few (3 – 5 %) *m/z* features were unique metabolites[11,12,15,31].

Chromatographic peak shape correlation of *m/z* features can identify such confounding effects in LC-based methods[19,21], but they are difficult to detect with FI-MS methods. Here, we used an experimental approach and examined significant *m/z* features in metabolite standards, which are most likely in-source derivates. Connecting these features via 51 mass differences of neutral losses, adducts, and isotopes described in

the literature[23] resulted in networks that explained the origin of 43 % of the significant features. MS2 spectra of the metabolite standards provided additional information about in-source fragmentation and explained another 6 % of the significant features.

Taken together, we found that FI-MS of single metabolites produces complex MS1 spectra, but they are explainable by known in-source modifications. The jury is out if in-source modifications are a bug or a feature for FI-MS data analysis: they may complicate or improve metabolite annotation. Therefore, the future challenge is to use FI-MS spectra of single metabolites to deconvolute FI-MS spectra from biological samples and, thereby, increase confidence of metabolite annotation. A first step is the construction of FI-MS databases with MS1 spectra of single metabolite standards to map in-source modifications across thousands of compounds. This is especially important since we found that the type of in-source modification depends on the metabolite classes (**Figure 48d**). Here, we provided a starting point with in-source modifications of 160 metabolites and mining MS1 spectra in existing databases like GNPS[32] and Metlin[33] may provide additional reference data.

7.9 Material and Methods

Chemicals and materials

Authentic metabolite standards were purchased from Merck KGaA (former Sigma-Aldrich, Germany). The standards were dissolved in water to a concentration of 1 mM if not stated otherwise. Standards were then further diluted with acetonitrile and methanol to a final concentration of 10 µM in 40:40:20 acetonitrile:methanol:water. The 10 µM metabolite standards were then added to an *E. coli* metabolite extract to yield a final concentration of 1 µM. *E. coli* cultures were in a M9 minimal medium, which contained: 22 mM KH_2PO_4, 42.2 mM Na_2HPO_4, 11.3 mM $(NH_4)2SO_4$, 8.56 mM NaCl, 100 µM $CaCl_2$ x 2 H_2O, 1 mM $MgSO_4$ x 7 H_2O, 60 µM $FeCl_3$, 6.3 µM $ZnSO_4$ x 7 H_2O, 7.6 µM $CoCl_2$ x 6 H_2O, 7.1 µM, 7 µM $CuCl_2$ x 2 H_2O, and $MnSO_4$ x 2 H_2O.

Metabolite extracts from *E. coli* cultures

5 mL LB medium was inoculated with *E. coli* MG1655 from a cryo stock. After 6 - 7 h of cultivation at 37°C, 10 µL of the culture was transferred to 5 mL M9 minimal medium

with 5 g/L glucose. For ^{13}C-labelled extracts, uniformly labelled ^{13}C-glucose was used (#CLM-1396, Cambridge Isotope Laboratories Inc., USA). The M9 precultures were grown overnight at 37°C and at 220 rpm shaking. 20 mL of M9 with 5 g/L ^{12}C- or ^{13}C-glucose was inoculated with the overnight culture to an optical density at 600 nm (OD) of 0.05. At an OD of 1, aliquots of 4 mL of the culture were vacuum-filtrated using 0.45 µm pore size filters (HVLP02500, Merck Millipore). The filters were transferred to -20°C cold 40:40:20 acetonitrile:methanol:water for metabolite extraction. After at least 30 min at -20°C, the metabolite extracts were centrifuged for 30 min at -9°C and 4,000 rpm. The supernatant was stored at -80°C.

Mass spectrometry

Samples were analyzed by FI-MS on an Agilent 6546 Series quadrupole time-of-flight mass spectrometer (Agilent Technologies, USA). The electrospray source was operated in negative and positive ionization mode. The mobile phase was 60:40 isopropanol:water buffered with 10 mM ammonium carbonate $(NH_4)_2CO_3$ and 0.04 % (v/v) ammonium hydroxide for both ionization modes, and the flow rate was 0.15 mL/min. For online mass axis correction, 2-propanol (in the mobile phase) and HP-921 were used for negative mode and purine and HP-921 were used for positive mode. Mass spectra were recorded in profile mode from 50 to 1100 *m/z* with a frequency of 1.4 spectra/s for 0.5 min using 10 Ghz resolving power. Source temperature was set to 225 °C, with 1 L/min drying gas and a nebulizer pressure of 20 psi. Fragmentor, skimmer, and octupole voltages were set to 120 V, 65 V, and 650 V, respectively.

Data preprocessing

Raw files were converted into "mzXML" format by MSConvert[26]. Further data processing was performed using MATLAB version R2021a (The Mathworks, Inc., USA). For each sample, an average spectrum was calculated from the ten scans with the highest ion counts. The spectra were resampled to 10^6 data points to align *m/z* values of all samples. Ion peaks were picked with the "findpeaks" function of MATLAB, using a peak height and prominence cutoff of 1,000 units. Hierarchical clustering with a tolerance of 7.5 mDa was used to bin peaks. For each peak bin, we calculated a centroid *m/z* value from the individual peak *m/z* values. Peaks were annotated to metabolites using the centroid *m/z*

value with a tolerance of 3 mDa. Z-scores were calculated from logarithmic mean signal intensities (triplicates). Z-scores above three were considered significant.

Calculation of mass differences

Mass differences between all significant features were calculated and combined for positive and negative ionization mode. Using the MATLAB function “histcounts”, all mass differences were assigned to one of 5 x 10^5 bins, which accounted for a mass resolution of ca. 2 mDa. The total number of mass differences in each bin is the frequency of a mass difference. In the resulting neutral loss spectrum (x-axis is the mass difference and y-axis the frequency), peaks were picked with 3 mDa tolerance using the “findpeaks” function with prominence and height cutoffs of 10 units. 51 peaks that matched mass differences in the literature[23] were then used for further analysis.

Construction of feature networks

Networks of significant *m/z* features were constructed for each spike-in sample in positive and negative ionization mode. Nodes are significant *m/z* features, and edges are putative modifications like adducts or chemical reactions, or isotopes. Edges are drawn between each pair of nodes if the mass difference between them matches the mass differences in the list of 51 explained mass differences. The fraction of explained features are calculated by counting all features that are either directly or indirectly connected to the (de-)protonated metabolite or fragment ion versus the total number of significant features for each spike-in standard. The feature networks were built with python v. 3.8.5 using the “networkx” toolbox. MS2 spectra were obtained from the Human Metabolome Database when experimental spectra were reported for a spike-in standard[27]. The experimental MS2 spectra from HMDB contained data that was acquired by high- and low-resolution mass spectrometers. Therefore, we matched our significant *m/z* features to the MS2 spectra with a tolerance of 100 mDa.

7.10 References

(1) Fuhrer, T.; Heer, D.; Begemann, B.; Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection-Time-of-Flight Mass Spectrometry. *Anal Chem* **2011**, *83* (18), 7074–7080. https://doi.org/10.1021/ac201267k.

(2) Beckmann, M.; Parker, D.; Enot, D. P.; Duval, E.; Draper, J. High-Throughput, Nontargeted Metabolite Fingerprinting Using Nominal Mass Flow Injection Electrospray Mass Spectrometry. *Nat Protoc* **2008**, *3* (3), 486–504. https://doi.org/10.1038/nprot.2007.500.

(3) Reiter, A.; Herbst, L.; Wiechert, W.; Oldiges, M. Need for Speed: Evaluation of Dilute and Shoot-Mass Spectrometry for Accelerated Metabolic Phenotyping in Bioprocess Development. *Anal Bioanal Chem* **2021**, *413* (12), 3253–3268. https://doi.org/10.1007/s00216-021-03261-3.

(4) Sarvin, B.; Lagziel, S.; Sarvin, N.; Mukha, D.; Kumar, P.; Aizenshtein, E.; Shlomi, T. Fast and Sensitive Flow-Injection Mass Spectrometry Metabolomics by Analyzing Sample-Specific Ion Distributions. *Nat Commun* **2020**, *11* (1), 3186. https://doi.org/10.1038/s41467-020-17026-6.

(5) Link, H.; Fuhrer, T.; Gerosa, L.; Zamboni, N.; Sauer, U. Real-Time Metabolome Profiling of the Metabolic Switch between Starvation and Growth. *Nature Methods* **2015**, *12* (11), 1091–1097. https://doi.org/10.1038/nmeth.3584.

(6) Fuhrer, T.; Zampieri, M.; Sévin, D. C.; Sauer, U.; Zamboni, N. Genomewide Landscape of Gene–Metabolome Associations in *Escherichia* Coli. *Molecular Systems Biology* **2017**, *13* (1), 907. https://doi.org/10.15252/msb.20167150.

(7) Anglada-Girotto, M.; Handschin, G.; Ortmayr, K.; Campos, A. I.; Gillet, L.; Manfredi, P.; Mulholland, C. V.; Berney, M.; Jenal, U.; Picotti, P.; Zampieri, M. Combining CRISPRi and Metabolomics for Functional Annotation of Compound Libraries. *Nat Chem Biol* **2022**, *18* (5), 482–491. https://doi.org/10.1038/s41589-022-00970-3.

(8) Holbrook-Smith, D.; Durot, S.; Sauer, U. High-throughput Metabolomics Predicts Drug–Target Relationships for Eukaryotic Proteins. *Molecular Systems Biology* **2022**, *18* (2). https://doi.org/10.15252/msb.202110767.

(9) Dubuis, S.; Ortmayr, K.; Zampieri, M. A Framework for Large-Scale Metabolome Drug Profiling Links Coenzyme A Metabolism to the Toxicity of Anti-Cancer Drug Dichloroacetate. *Communications Biology* **2018**, *1* (1), 101. https://doi.org/10.1038/s42003-018-0111-x.

(10) Rathahao-Paris, E.; Alves, S.; Boussaid, N.; Picard-Hagen, N.; Gayrard, V.; Toutain, P.-L.; Tabet, J.-C.; Rutledge, D. N.; Paris, A. Evaluation and Validation of an Analytical Approach for High-Throughput Metabolomic Fingerprinting Using Direct Introduction–High-Resolution Mass Spectrometry: Applicability to Classification of Urine of Scrapie-Infected Ewes. *Eur J Mass Spectrom (Chichester)* **2019**, *25* (2), 251–258. https://doi.org/10.1177/1469066718806450.

(11) Hartl, J.; Kiefer, P.; Kaczmarczyk, A.; Mittelviefhaus, M.; Meyer, F.; Vonderach, T.; Hattendorf, B.; Jenal, U.; Vorholt, J. A. Untargeted Metabolomics Links Glutathione to Bacterial Cell Cycle Progression. *Nat Metab* **2020**, *2* (2), 153–166. https://doi.org/10.1038/s42255-019-0166-0.

(12) Wang, L.; Xing, X.; Chen, L.; Yang, L.; Su, X.; Rabitz, H.; Lu, W.; Rabinowitz, J. D. Peak Annotation and Verification Engine for Untargeted LC–MS Metabolomics. *Anal. Chem.* **2019**, *91* (3), 1838–1846. https://doi.org/10.1021/acs.analchem.8b03132.

(13) Kachman, M.; Habra, H.; Duren, W.; Wigginton, J.; Sajjakulnukit, P.; Michailidis, G.; Burant, C.; Karnovsky, A. Deep Annotation of Untargeted LC-MS Metabolomics Data with Binner. *Bioinformatics* **2020**, *36* (6), 1801–1806. https://doi.org/10.1093/bioinformatics/btz798.

(14) Sindelar, M.; Patti, G. J. Chemical Discovery in the Era of Metabolomics. *J. Am. Chem. Soc.* **2020**, *142* (20), 9097–9105. https://doi.org/10.1021/jacs.9b13198.

(15) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **2017**, *89* (19), 10397–10406. https://doi.org/10.1021/acs.analchem.7b02380.

(16) Xue, J.; Domingo-Almenara, X.; Guijas, C.; Palermo, A.; Rinschen, M. M.; Isbell, J.; Benton, H. P.; Siuzdak, G. Enhanced In-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification. *Anal Chem* **2020**, *92* (8), 6051–6059. https://doi.org/10.1021/acs.analchem.0c00409.

(17) Purwaha, P.; Silva, L. P.; Hawke, D. H.; Weinstein, J. N.; Lorenzi, P. L. An Artifact in LC-MS/MS Measurement of Glutamine and Glutamic Acid: In-Source Cyclization to Pyroglutamic Acid. *Anal. Chem.* **2014**, *86* (12), 5633–5637. https://doi.org/10.1021/ac501451v.

(18) Guo, J.; Shen, S.; Xing, S.; Yu, H.; Huan, T. ISFrag: De Novo Recognition of In-Source Fragments for Liquid Chromatography–Mass Spectrometry Data. *Anal. Chem.* **2021**, *93* (29), 10243–10250. https://doi.org/10.1021/acs.analchem.1c01644.

(19) Schmid, R.; et al. Ion Identity Molecular Networking for Mass Spectrometry-Based Metabolomics in the GNPS Environment. *Nat Commun* **2021**, *12* (1), 3832. https://doi.org/10.1038/s41467-021-23953-9.

(20) Senan, O.; Aguilar-Mogas, A.; Navarro, M.; Capellades, J.; Noon, L.; Burks, D.; Yanes, O.; Guimerà, R.; Sales-Pardo, M. CliqueMS: A Computational Tool for Annotating in-Source Metabolite Ions from LC-MS Untargeted Metabolomics Data Based on a Coelution Similarity Network. *Bioinformatics* **2019**, *35* (20), 4089–4097. https://doi.org/10.1093/bioinformatics/btz207.

(21) Xu, Y.-F.; Lu, W.; Rabinowitz, J. D. Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography–Mass Spectrometry-Based Metabolomics. *Analytical Chemistry* **2015**, *87* (4), 2273–2281. https://doi.org/10.1021/ac504118y.

(22) Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M. A.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; Böcker, S. Systematic Classification of Unknown Metabolites Using High-Resolution Fragmentation Mass Spectra. *Nat Biotechnol* **2021**, *39* (4), 462–471. https://doi.org/10.1038/s41587-020-0740-8.

(23) Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A. D.; Shen, Y.; Cowan, A.; McReynolds, M. R.; Kennedy, B. J.; Lato, A. M.; Campagna, S. R.; Singh, M.; Rabinowitz, J. D. Metabolite Discovery through Global Annotation of Untargeted Metabolomics Data. *Nat Methods* **2021**, *18* (11), 1377–1385. https://doi.org/10.1038/s41592-021-01303-3.

(24) Nothias, L.-F.; et al. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat Methods* **2020**, *17* (9), 905–908. https://doi.org/10.1038/s41592-020-0933-6.

(25) Stricker, T.; Bonner, R.; Lisacek, F.; Hopfgartner, G. Adduct Annotation in Liquid Chromatography/High-Resolution Mass Spectrometry to Enhance Compound Identification. *Anal Bioanal Chem* **2021**, *413* (2), 503–517. https://doi.org/10.1007/s00216-020-03019-3.

(26) Chambers, M. C.; et al. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nature Biotechnology* **2012**, *30* (10), 918–920. https://doi.org/10.1038/nbt.2377.

(27) Wishart, D. S.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res* **2022**, *50* (D1), D622–D631. https://doi.org/10.1093/nar/gkab1062.

(28) Guder, J. C.; Schramm, T.; Sander, T.; Link, H. Time-Optimized Isotope Ratio LC–MS/MS for High-Throughput Quantification of Primary Metabolites. *Anal. Chem.* **2017**, *89* (3), 1624–1631. https://doi.org/10.1021/acs.analchem.6b03731.

(29) Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal Chem* **2021**, *93* (34), 11692–11700. https://doi.org/10.1021/acs.analchem.1c01465.

(30) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; Feist, A. M.; Palsson, B. O. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat Biotechnol* **2017**, *35* (10), 904–908. https://doi.org/10.1038/nbt.3956.

(31) Mahieu, N. G.; Spalding, J. L.; Gelman, S. J.; Patti, G. J. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.Unity Algorithm. *Anal. Chem.* **2016**, *88* (18), 9037–9046. https://doi.org/10.1021/acs.analchem.6b01702.

(32) Wang, M. et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(33) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan, D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**, *90* (5), 3156–3164. https://doi.org/10.1021/acs.analchem.7b04424.

Chapter 8 Acquisition of unknown MS2 fragments for low abundant and unavailable metabolites with CRISPR interference

Niklas Farke, Andreas Verhuelsdonk, Hannes Link

8.1 Abstract

To identify metabolites, mass spectrometry-based metabolomics approaches rely on accurate spectral databases that contain fragment signatures of all metabolites. These signatures are achieved by fragmenting metabolites in the mass spectrometer by collision induced dissociation. The combination of retention time, precursor mass, fragment masses and their intensities are usually used for metabolite identification. However, while this is the standard approach for most metabolites, there are certain metabolites for which this information is not available, because they are too complex for chemical synthesis, or too low abundant to produce biologically. Consequently, these metabolites cannot be verified with MS2 information. This is the case for many metabolites in purine, histidine and isoprenoid metabolism in *E. coli*. Only 6/31 of the pathway intermediates were available on Sigma Aldrich and for only 7/31 metabolites there was experimental MS2 information available on the human metabolome database. Here, we used CRISPR interference to enrich low-abundant pathway intermediates in purine, histidine, and isoprenoid biosynthesis. First, we used FI-MS to measure the samples and we observed specific increases in 21/25 mass peaks. Finally, we used an LC-MS/MS approach to generate MS2 spectra for: adenylo succinate (DCAMP) and 2-(formamido)-N-(5-phospho-β-D-ribosyl)-acetamide (FGAM).

8.2 Introduction

In tandem mass spectrometry-based metabolomics, metabolites are first separated by liquid chromatography based on their physicochemical properties[1–3]. Metabolites are then ionized via electrospray ionization (ESI[4]) and fragmented in the mass spectrometer by collision induced dissociation (CID[5]). Metabolites are commonly identified by their retention times, precursor masses, and fragmentation patterns. Measured mass spectra are compared to spectral databases that contain MS2 information for a large variety of metabolites[1–3,6]. Common databases are the Metabolite and Chemical Entity Database (METLIN)[7], and the Human Metabolome Database (HMDB[8]). For example, METLIN contains MS2 data for roughly 10 million compounds. However, this is only 1 % of the

93 million compounds listed in PubChem, posing the question why there is no MS2 information for the remaining 99% of compounds[7].

Experimental approaches commonly use authentic metabolite standards to perform MS2 scans like multiple reaction monitoring to measure fragment ions[3]. Metabolite standards are either synthesized chemically or produced biotechnologically[9]. This can require complicated approaches to synthesize some metabolites, as shown for intermediates in the methylerythritol phosphate pathway[10]. Some metabolites, however, may be too complex to synthesize chemically or they are simply too low abundant in microbial metabolism, causing a knowledge gap in MS2 databases[7].

To understand metabolism in *E. coli*, it is crucial to measure all metabolites under various conditions. This information is then integrated with mathematical models or data driven approaches to infer biological mechanisms[11–13]. Incomplete databases prohibit confident annotation of these metabolites and while several prediction tools have been developed to work around these limitations, predictions may not always be accurate and need to be validated experimentally[14–17]. Histidine, purine and isoprenoid biosynthesis in *E. coli* are three pathways in *E. coli* primary metabolism in which many metabolites lack experimental MS2 information[8]. Consequently, confident annotation of several metabolites in these pathways is still problematic.

Here, we used CRISPR interference (CRISPRi)[18] to enrich low-abundant pathway intermediates in purine, histidine, and isoprenoid biosynthesis. First, we used flow-injection mass spectrometry (FI-MS[19,20]) to measure specific increases in 21/25 mass peaks. Finally, we used an LC-MS/MS approach to generate MS2 spectra for two example metabolites: DCAMP (adenylo succinate) and FGAM (2-(formamido)-N-(5-phospho-β-D-ribosyl)-acetamide). For FGAM, we discovered a novel MS2 fragment that, to our knowledge, was neither predicted, nor measured yet.

8.3 FI-MS identifies significant metabolome changes of pathway intermediates in purine, histidine and isoprenoid biosynthesis

Histidine biosynthesis, purine biosynthesis and the methylerythritol phosphate (MEP) pathway are important anabolic pathways in *E. coli* and essential for growth on glucose

minimal medium. Despite their importance, there is little MS2 information available for most metabolites because they are commercially not available. This is because they are low abundant and difficult to synthesize chemically. As a starting point of our analysis, we selected CRISPRi strains from our genome wide CRISPRi library (**Chapter 6**) to knockdown genes in these three pathways. All three pathways are linear, and histidine and purine biosynthesis are connected via the enzyme HisH/F that produces AICAR from PRLP (**Figure 50**).

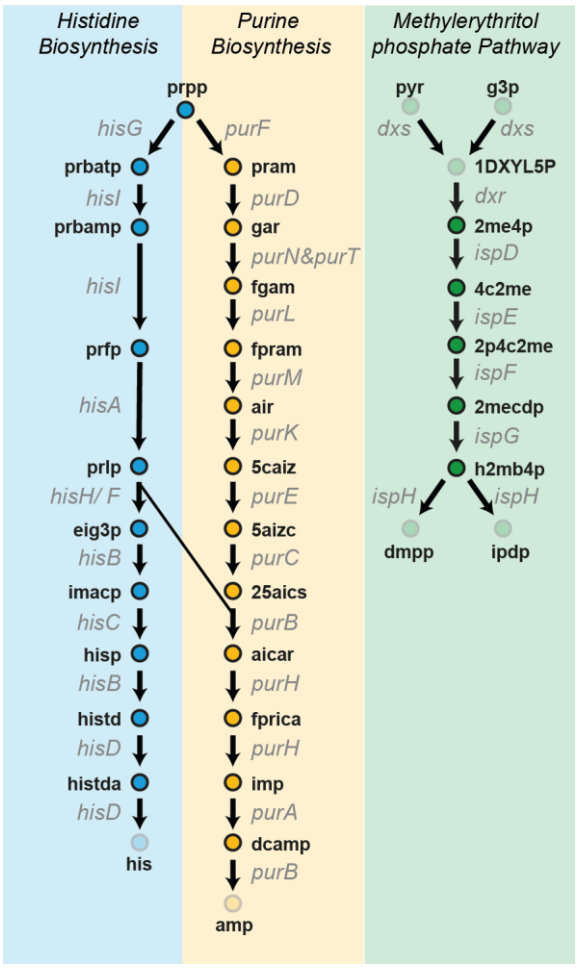


Figure 50. Histidine, Purine, and Isoprenoid metabolism*. Metabolic map of histidine biosynthesis (blue), purine biosynthesis (orange), and the methylerythritol phosphate (MEP) pathway (green). Arrows are metabolic reactions and spheres are metabolites. Transparent metabolites were not considered in the analysis.*

In total, we targeted eight genes in histidine biosynthesis (*hisA*, *hisB*, *hisC*, *hisD*, *hisF*, *hisG*, *hisH*, *hisI*), twelve genes in purine biosynthesis (*purA*, *purB*, *purC*, *purD*, *purE*, *purF*, *purH*, *purK*, *purl*, *purM*, *purN*, *purT*), and five genes in the MEP pathway (*ispD*, *ispE*, *ispF*, *ispG*, *ispH*) (**Appendix, Table 5**). We then cultured these strains in LB medium, induced

CRISPRi with aTC, and then measured the metabolome of each strain with FI-MS (n = 1). To identify significant (z-score > 3) metabolome changes, we inspected the ion peaks of the (de-)protonated metabolites in the MS1 spectrum (**Figure 51**). We observed increases in 21/25 ion peaks (HISTDA, PRBATP, and FPRICA had no significantly changed ion peaks, and HISTD had a significantly changed ion peak which was below the intensity cutoff of 1000). Overall, each of the 21 metabolites accumulated in at least one condition (**Figure 51**).

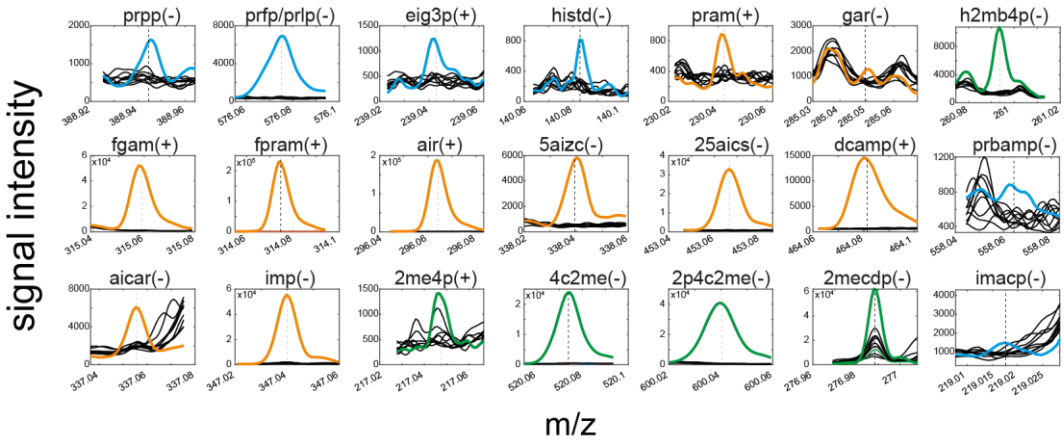


Figure 51. MS1 spectra for 21 metabolites in the purine(orange), histidine (blue), and MEP pathway (green). *The bracket indicates if the MS1 spectrum was recorded in positive (+) or negative (-) ionization mode. The coloured spectra are CRISPRi strains of significantly changed ion peaks. The black spectra are CRISPRi strains that did not result in a significantly changed ion peak.*

For example, the substrate of *ispH* is H2MB4P. H2MB4P accumulated in the *ispH* strain (green), while the other samples only had a baseline signal at the same *m/z* value in the MS1 spectrum. Similarly, 5AIZC is the substrate of *purC* and accumulated only in the *purC* strain, while the other samples only had a baseline signal at the same *m/z* value. Next, we investigated the specificity of each CIRPSRi knockdown (**Figure 52**). 16/21 metabolites accumulated as the substrate of a CRISPRi target. However, the specificity of the perturbation varied between the different pathways. For example, in the histidine pathway, many bottlenecks (*hisD, hisC, hisB*) seemed to cause a tailback leading to accumulation of upstream metabolites. To a lesser extent, we observed the same effect in the purine and MEP pathways. Moreover, a bottleneck in PurC caused the accumulation of AICAR, which is the product of the reaction. In this case, the bottleneck likely causes redirection of flux through the histidine pathway, bypassing the bottleneck

and leading to an accumulation of AICAR. However, control mechanisms behind this effect are unknown.

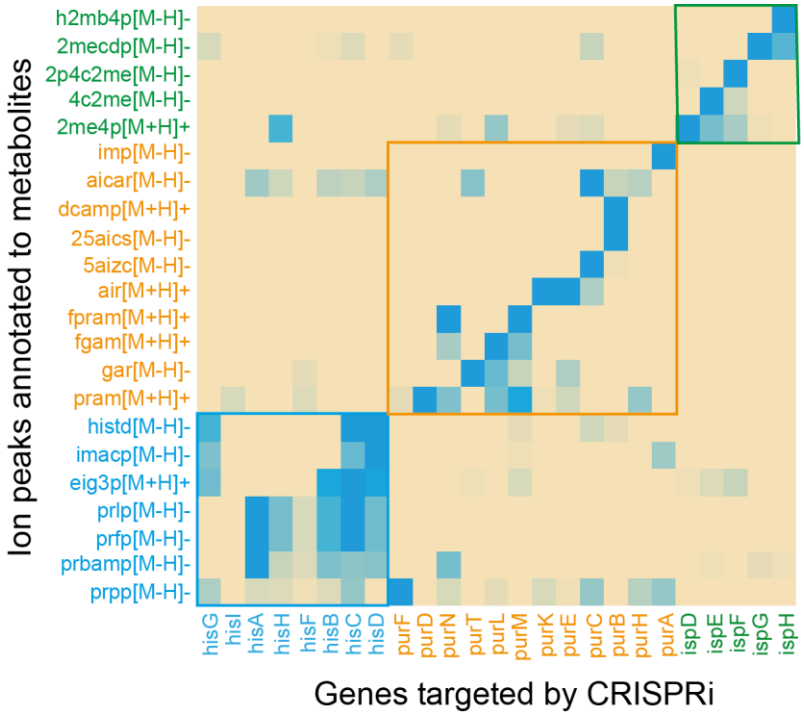


Figure 52. Heatmap of CRISPRi knockdowns of targets in histidine, purine, and isoprenoid biosynthesis. *Heatmap of all CRISPRi target genes and ion peaks that were annotated to metabolites in either positive or negative ionization mode. Z-scores are shown for all ion peaks of all CRISPRi strains. Orange indicates a z-score <= 0. A blue square indicates are z-score > 0. A blue box indicates genes and ion peaks in the histidine pathway, an orange box indicates genes and ion peaks in the purine pathway, and a green box indicates genes and ion peaks in the MEP pathway.*

In conclusion, we used CRISPRi to knockdown genes in purine, histidine, and MEP biosynthesis. We then used FI-MS to identify peaks of 21 metabolites, which otherwise would only show a baseline signal. In total, 16/21 metabolites were the substrates of the genes targeted by CRISPRi. Here, we used CRISPRi to specifically accumulate low abundant metabolites. In the next step, we sought to acquire MS2 fragments for some of these metabolites.

8.4 Identification of MS2 fragments for metabolites in histidine, purine, and isoprenoid metabolism with an LC-MS/MS approach

To measure MS2 spectra of the respective substrates of the CRISPRi strains, we used LC-MS/MS with a triple quadrupole mass spectrometer. In contrast to FI-MS, metabolites

are first separated by liquid chromatography. The first quadrupole filters for specific precursor ions based on their *m/z* value. In the second quadrupole, selected ions are fragmented by CID. The third quadrupole then filters for fragment masses. Here, we used a product scan approach to record MS2 spectra for the substrates of the CRISPRi strains. In a product scan, pre-defined ion masses (here the substrates) are filtered out and fragmented. The result is a MS2 spectrum of all isolated ions in the defined *m/z*-window.

We tested this approach for *purB* and *purM*. The substrate of *purB* is DCAMP and the substrate of *purM* is FGAM. While there are experimental MS2 spectra available for DCAMP (adenylo-succinate), only theoretical MS2 spectra were available for FGAM (2-(formamido)-N-(5-phospho-β-D-ribosyl)-acetamide). We used the induced CRISPRi strains, and the controls were the uninduced CRISPRi strains. While all uninduced CRISPRi strains also have a basal CRISPRi activity, we expected the difference between induced and uninduced to be significant.

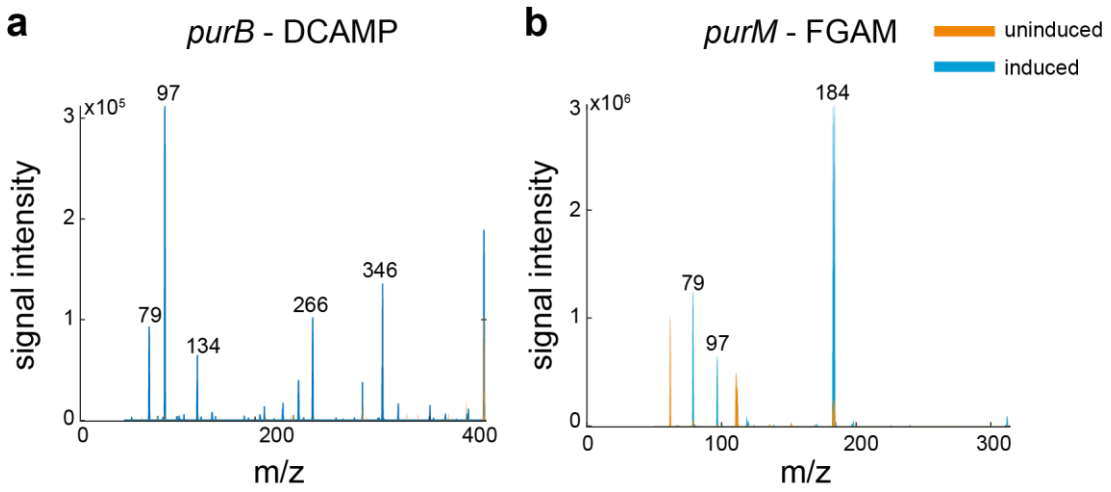


***Figure 53**. **MS2 from product scans**. MS2 spectra of **a)** DCAMP (purB CRISPRi strain) and **b)** FGAM (purM CRISPRi strain). Blue spectra are the induced strains and orange spectra are the uninduced strains. Numbers above selected peaks indicate the m/z of the peak.*

The recorded MS2 spectra showed a clear difference between induced CRISPRi strains and uninduced strains (**Figure 53**). In the MS2 spectrum of the *purB* strain (**Figure 53a**), we selected the five highest peaks and compared them against experimental MS2 spectra of DCAMP on HMDB. In total, all selected fragment *m/z* matched reported peaks

on HMDB (LC-ESI, negative mode: *m/z* = 79, *m/z* = 97, *m/z* = 134, *m/z* = 266, *m/z* = 346), thus validating our approach.

Next, we analyzed the *purM* strain to obtain MS2 spectra for FGAM. In the MS2 spectrum of the *purM* strain, we selected three peaks (*m/z* = 79, *m/z* = 97, *m/z* = 184) (**Figure 53b**). While there were no experimental spectra available, we compared the selected peaks to predicted MS2 spectra of FGAM. In total, *m/z* = 79 and *m/z* = 97 matched the predicted spectrum. However, the *m/z* = 184 fragment ion was not reported on HMDB. These results indicate that this approach can be used to acquire MS2 spectra of low abundant metabolites that were accumulated via CRISPRi.

8.5 Discussion

Databases with tandem mass spectral data are important to unequivocally identify metabolites in LC-MS/MS approaches. Commonly, authentic standards are measured and fragmented to record MS2 fragment ions. However, many metabolites are either too complex to synthesize or too low abundant to produce biotechnologically. Thus, this knowledge gap limits LC-MS/MS approaches to focus on available metabolites.

Here, we used CRISPRi to enrich low-abundant metabolites in histine, purine, and isoprenoid biosynthesis. We showed that CRISPRi-based bottlenecks can cause accumulation of substrate metabolites and that these metabolites can then be measured by FI-MS. Using an LC-MS/MS product scan approach, we identified known MS2 fragment ions for DCAMP and an unknown fragment ion for FGAM. Our approach is easier compared to approaches that first have to purify metabolic enzymes for subsequent *in-vitro* synthesis of metabolites[10]. Moreover, CRISPRi can be easily applied to other organisms and therefore this approach has the potential to produce MS2 fragments for metabolites of various chemical classes in many organisms.

These results indicate the potential of CRISPRi screens to complement current MS2 databases like METLIN[7] or HMDB[8]. Although computational prediction tools have been successfully used to predict fragment spectra[14,17], predicted spectra need to be validated by experimental data. MS2 spectra can be misannotated due to false positive MS1 ions, or due to isomers/isobars that share the same retention times. Thus, to

improve confidence in the MS2 spectra, an isotope ratio approach should be used to validate the MS2 information in the future. Moreover, high-resolution mass spectrometers should be used to achieve high resolution MS2 spectra.

8.6 Methods

Flow-injection metabolomics (FI-MS) and data processing

Samples were analysed by FI-MS on an Agilent 6546 Series quadrupole time-of-flight mass spectrometer (Agilent Technologies, USA). The electrospray source was operated in negative and positive ionization mode. The mobile phase was 60:40 isopropanol:water buffered with 10 mM ammonium carbonate $(NH_4)_2CO_3$ and 0.04 % (v/v) ammonium hydroxide for both ionization modes, and the flow rate was 0.15 mL/min. For online mass axis correction, 2-propnaol (in the mobile phase) and HP-921 were used for negative mode and purine and HP-921 were used for positive mode. Mass spectra were recorded in profile mode from 50 to 1,100 m/z with a frequency of 1.4 spectra/s for 0.5 min using 10 Ghz resolving power. Source temperature was set to 225 °C, with 1 l/min drying gas and a nebulizer pressure of 20 psi. Fragmentor, skimmer, and octupole voltages were set to 120 V, 65 V, and 650 V, respectively.

Raw files were converted into “mzXML” format by MSConvert. Further data processing was performed using MATLAB version R2021a (The Mathworks, Inc., USA). For each sample, an average spectrum was calculated from the ten scans with the highest ion counts. The spectra were resampled to 10^6 data points to align m/z values of all samples. Ion peaks were picked with the “findpeaks” function of MATLAB, using a peak height and prominence cutoff of 750 units. Hierarchical clustering with a tolerance of 7.5 mDa was used to bin peaks. For each peak bin, we calculated a centroid m/z value from the individual peak m/z values. Peaks were annotated to metabolites using the centroid m/z value with a tolerance of 3 mDa.

LC-MS/MS with a triple quadrupole mass spectrometer and data processing

An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30 °C. The injection volume was 3 µl. LC solvent A was water with 10 mM ammonium formate and 0.1 % formic acid (v/v)

for acidic conditions ans water with 10 mM ammonium carbonate and 0.2 % ammonium hydroxide for basic conditions. LC solvents B were acetonitrile with 0.1 % formic acid (v/v) for acidic conditions and acetonitrile without additive for basic conditions.

An Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) was used for the product ion scans. Source gas temperature was set to 200 °C, with 14 L/min drying gas and a nebulizer pressure of 24 psi. Sheath gas temperature was set to 300 °C and flow to 11 l/min. Electrospray nozzle and capillary voltages were set to 500 and 2500 V. Scan time was set to 500 ms, Fragmentor was set to 380 V, cell accelerator voltage was set to 5, and step size was 0.1. Product scans were made for collision energies of 10 V, 25V, and 40 V. For the purB strain, masses were scanned from *m/z* = 50 to *m/z* = 464.

Raw files were converted into “mzXML” format by MSConvert. Further data processing was performed using MATLAB version R2021a (The Mathworks, Inc., USA). For each sample, intensities were summed up over all retention times to achieve a 2-dimensional MS1 spectrum.

8.7 References

(1) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin H, C.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat Methods* **2020**, *17* (9), 905–908. https://doi.org/10.1038/s41592-020-0933-6.

(2) Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A. D.; Shen, Y.; Cowan, A.; McReynolds, M. R.; Kennedy, B. J.; Lato, A. M.; Campagna, S. R.; Singh, M.; Rabinowitz, J. D. Metabolite Discovery through Global Annotation of Untargeted Metabolomics Data. *Nat Methods* **2021**, *18* (11), 1377–1385. https://doi.org/10.1038/s41592-021-01303-3.

(3) Guder, J. C.; Schramm, T.; Sander, T.; Link, H. Time-Optimized Isotope Ratio LC-MS/MS for High-Throughput Quantification of Primary Metabolites. *Anal Chem* **2017**, *89* (3), 1624–1631. https://doi.org/10.1021/acs.analchem.6b03731.

(4) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459. https://doi.org/10.1021/j150664a002.
(5) Wells, J. M.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol* **2005**, *402*, 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.
(6) Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; Korf, A.; Pluskal, T.; Kameník, Z.; Jarmusch, A. K.; Caraballo-Rodríguez, A. M.; Weldon, K. C.; Nothias-Esposito, M.; Aksenov, A. A.; Bauermeister, A.; Albarracin Orio, A.; Grundmann, C. O.; Vargas, F.; Koester, I.; Gauglitz, J. M.; Gentry, E. C.; Hövelmann, Y.; Kalinina, S. A.; Pendergraft, M. A.; Panitchpakdi, M.; Tehan, R.; Le Gouellec, A.; Aleti, G.; Mannochio Russo, H.; Arndt, B.; Hübner, F.; Hayen, H.; Zhi, H.; Raffatellu, M.; Prather, K. A.; Aluwihare, L. I.; Böcker, S.; McPhail, K. L.; Humpf, H.-U.; Karst, U.; Dorrestein, P. C. Ion Identity Molecular Networking for Mass Spectrometry-Based Metabolomics in the GNPS Environment. *Nat Commun* **2021**, *12* (1), 3832. https://doi.org/10.1038/s41467-021-23953-9.
(7) Xue, J.; Guijas, C.; Benton, H. P.; Warth, B.; Siuzdak, G. METLIN MS2 Molecular Standards Database: A Broad Chemical and Biological Resource. *Nat Methods* **2020**, *17* (10), 953–954. https://doi.org/10.1038/s41592-020-0942-5.
(8) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res* **2007**, *35* (Database issue), D521-526. https://doi.org/10.1093/nar/gkl923.
(9) Mitsuhashi, S. Current Topics in the Biotechnological Production of Essential Amino Acids, Functional Amino Acids, and Dipeptides. *Current Opinion in Biotechnology* **2014**, *26*, 38–44. https://doi.org/10.1016/j.copbio.2013.08.020.
(10) Li, Z.; Sharkey, T. D. Metabolic Profiling of the Methylerythritol Phosphate Pathway Reveals the Source of Post-Illumination Isoprene Burst from Leaves. *Plant, Cell & Environment* **2013**, *36* (2), 429–437. https://doi.org/10.1111/j.1365-3040.2012.02584.x.
(11) Gerosa, L.; Haverkorn van Rijsewijk, B. R. B.; Christodoulou, D.; Kochanowski, K.; Schmidt, T. S. B.; Noor, E.; Sauer, U. Pseudo-Transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Syst* **2015**, *1* (4), 270–282. https://doi.org/10.1016/j.cels.2015.09.008.
(12) Orth, J. D.; Thiele, I.; Palsson, B. Ø. What Is Flux Balance Analysis? *Nat Biotechnol* **2010**, *28* (3), 245–248. https://doi.org/10.1038/nbt.1614.
(13) Kochanowski, K.; Volkmer, B.; Gerosa, L.; Haverkorn van Rijsewijk, B. R.; Schmidt, A.; Heinemann, M. Functioning of a Metabolic Flux Sensor in Escherichia Coli. *Proc Natl Acad Sci U S A* **2013**, *110* (3), 1130–1135. https://doi.org/10.1073/pnas.1202582110.

(14) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID. *Proc Natl Acad Sci U S A* **2015**, *112* (41), 12580–12585. https://doi.org/10.1073/pnas.1509788112.
(15) Böcker, S. Searching Molecular Structure Databases Using Tandem MS Data: Are We There Yet? *Curr Opin Chem Biol* **2017**, *36*, 1–6. https://doi.org/10.1016/j.cbpa.2016.12.010.
(16) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8* (2), E31. https://doi.org/10.3390/metabo8020031.
(17) Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal Chem* **2021**, *93* (34), 11692–11700. https://doi.org/10.1021/acs.analchem.1c01465.
(18) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152* (5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022.
(19) Sarvin, B.; Lagziel, S.; Sarvin, N.; Mukha, D.; Kumar, P.; Aizenshtein, E.; Shlomi, T. Fast and Sensitive Flow-Injection Mass Spectrometry Metabolomics by Analyzing Sample-Specific Ion Distributions. *Nat Commun* **2020**, *11* (1), 3186. https://doi.org/10.1038/s41467-020-17026-6.
(20) Fuhrer, T.; Heer, D.; Begemann, B.; Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection-Time-of-Flight Mass Spectrometry. *Anal Chem* **2011**, *83* (18), 7074–7080. https://doi.org/10.1021/ac201267k.

Chapter 9 Key findings

To fully understand metabolism, comprehensive multi-omics data need to be interpreted with the help of mathematical models. In **Chapter 2** – **Chapter 5**, we developed small mechanistic mathematical models of *E. coli* metabolism that are regulated allosterically, transcriptionally, or by both mechanisms. One of the key topics of this thesis was the interplay of metabolism and enzyme-level regulation. We demonstrated that considering both layers in mathematical models is crucial to achieve a profound understanding of metabolism, and that this knowledge can aid the design of better production strains in biotechnology. By combining ensemble modelling and robustness analysis techniques, we described the functional implications of feedback regulation on metabolism. We showed that feedback regulation can increase robustness of a system, but that it can also cause system failure under certain circumstances. Moreover, we focused on small mechanistic models because they are easier to analyse, and they enable qualitative predictions that are difficult to achieve with bigger models due to epistemic uncertainties.

To overcome some of these uncertainties, we mapped transcriptional regulation of E. coli metabolism in **Chapter 6**. We used a metabolism wide CRISPRi library to perturb gene expression of 283 metabolic genes and measured the proteome and the metabolome in each condition. A pathway-based analysis of the proteome data then revealed transcriptional feedback regulation in 24 metabolic pathways. Moreover, we recovered known metabolite-TF interactions in arginine and methionine biosynthesis and suggested a putative metabolite-TF interaction in NAD *de novo* biosynthesis.

Finally, a profound understanding of metabolism requires reliable multi-omics data to complement mathematical models, or to infer functional metabolite-protein interactions in a data-driven way. However, technological limitations and the complexity of the metabolome prohibits us from measuring the complete metabolome with confidence. To solve technological limitations, we first need to be aware of them. Thus, in **Chapter 7**, we evaluated the flow-injection mass spectrometry (FI-MS) approach. We showed that reactions taking place in the ion source cause FI-MS to produce complex MS1 spectra, leading to misannotations. To confirm metabolite identities, LC-MS/MS

approaches commonly rely on reference databases that contain metabolite fragment information. However, for many metabolites this information is currently unavailable because they are too low abundant to measure, or too complex to synthesize. In **Chapter 8**, we therefore developed an approach that uses CRISPRi to enrich low-abundant metabolites and measure their MS2 fragments. This approach has the potential to complement MS2 databases.

In the following, key findings of each chapter are highlighted in detail:

In **Chapter 2** we studied the interplay between allosteric regulation and transcriptional regulation in amino acid biosynthesis. We developed a mechanistic mathematical model of amino acid biosynthesis that combined allosteric regulation of enzyme activity and transcriptional regulation of enzyme abundance. The model showed that allosteric feedback and transcriptional feedback act in concert to finely tune amino acid concentrations to balance robustness and efficiency in amino acid biosynthesis. We showed that an optimal balance between both objectives is reached when both feedbacks are equally strong. Moreover, equally strong feedback reduces the likelihood of amino acid oscillations that can occur if allosteric feedback is weak. We verified the model predictions using multi-omics data of allosterically and transcriptionally dysregulated *E. coli* mutants.

In **Chapter 3**, we studied metabolic burden of engineered glycerol-producing *E. coli* with a mathematical model of glycolysis that combined metabolism and transcriptional regulation via the transcription factor Cra. Induction of the glycerol pathway in *E. coli* caused growth defects and low glycerol titers. Analysis of the mathematical model suggested that the reason for this behaviour is transcriptional misregulation of glycolysis by the transcription factor Cra. Induction of glycerol production caused a depletion of fructose 1,6-bisphosphate, a lower glycolysis flux, and a low growth rate. Subsequently, fructose 1,6-bisphosphate activates the transcription factor Cra, which amplifies these effects and causes system failure. A robustness analysis of the model suggested that putting the glycerol pathway under transcriptional control of Cra achieves a more robust production strain with higher glycerol titers. We verified the model predictions experimentally by engineering Cra regulation in the glycerol pathway. This way, we

created a competitive glycerol production strain that maintained stable growth at high induction levels.

In **Chapter 4**, we studied how allosteric activation of CarAB by ornithine confers robustness of the branch point between arginine and pyrimidine biosynthesis. A CRISPRi knockdown of CarAB caused ornithine to accumulate 512-fold. To study the implications of the allosteric activation on the branch point between arginine biosynthesis and pyrimidine biosynthesis, we developed a mechanistic mathematical model. The model showed that allosteric activation of CarAB by ornithine buffers the knockdown of CarAB, alleviating a downstream bottleneck at the branch point and, subsequently, stabilizing metabolite end products in arginine and pyrimidine biosynthesis.

In **Chapter 5**, we developed a mechanistic mathematical model of glycolysis to study the causes of pyruvate oscillations. The model consisted of the phosphotransferase system and three of the most relevant allosteric regulations of glycolysis. We simulated 2,000 parameter sets based on literature parameter values. Decreasing the glucose uptake rate of the model by 5 % caused sustained oscillations in 20 % of the parameter sets. Increasing the glucose uptake rate 10-fold caused oscillations in 2 % of the parameter sets. On average, the period of the oscillations was three minutes. The model predicted that high cooperativity and high saturation of the pyruvate dehydrogenase, as well as the feed forward activation of the pyruvate kinase by fructose 1,6-bisphosphate as the main causes for sustained oscillations of pyruvate levels.

In **Chapter 6**, we performed CRISPRi knockdowns of 283 metabolic genes in *E. coli* and measured the proteome and metabolome in these strains to map transcriptional feedback regulation. The proteome data showed that 45 % of the target genes were significantly downregulated (z-score < -3) and that, on average, each knockdown caused 40 other proteins to change significantly as part of a compensatory response. Analysis of the proteome data then revealed transcriptional feedback regulation in 24 metabolic pathways of *E. coli*, as well as potential unknown transcriptional regulation of the genes *menH* and *folM*. Using FI-MS, we identified accumulation of substrates in 66 CRISPRi strains. Using the metabolome data, we recovered known metabolite regulators of the transcription factors ArgR and MetR. Finally, we inferred a putative functional interaction between nicotinate and NadR.

Chapter 7. To evaluate FI-MS, we spiked 160 authentic standards of *E. coli* primary metabolism into *E. coli* metabolite extracts and measured the 160 spike-in samples by flow-injection metabolomics. We annotated 154/160 metabolites in their single protonated or single deprotonated form, and 134 of the ion peaks increased in the corresponding spike-in standard. Besides these expected increases of ion peaks, we identified significant increases of 11,013 ion peaks in positive and negative mode combined across all spike-in samples, which is on average 68 additional *m/z* features for each spike-in sample. We then used known mass differences and MS2 information to connect the *m/z* features with a network approach. This way, we could explain on average 49 % of all significantly changed ion peaks. Finally, our results indicated that on average one true increase of a single metabolite will cause one (false) increase of an ion peak that is misannotated to another metabolite. Thus, flow-injection metabolomics is prone to false positives.

In **Chapter 8**, we developed an approach to generate MS2 fragments from low abundant and commercially unavailable, or expensive metabolites of *E. coli* metabolism. Therefore, we used CRISPRi to knockdown genes in histidine, purine, and isoprenoid biosynthesis pathways. This way, we systematically enriched the substrates of the CRISPRi target genes. Using flow-injection metabolomics we measured accumulations in 21/25 pathway metabolites. We then used targeted liquid chromatography tandem mass spectrometry to generate MS2 spectra for DCAMP and FGAM. For DCAMP, we recovered known MS2 fragments and for FGAM, we discovered an unknown fragment, thus validating this approach.

Chapter 10 Conclusions and Outlook

In the first part of this thesis (**Chapter 2** – **Chapter 5**), we developed small mechanistic mathematical models to study feedback control in metabolic networks. We observed that feedback control often renders the system robust against perturbations, but that under certain circumstances, the same regulation mechanisms can cause system failure. This was especially apparent in **Chapter 3**, where Cra regulation of glycolysis caused system failure after inducing an engineered glycerol pathway. Similarly, in **Chapter 5** we showed that feed forward activation of the pyruvate kinase by fructose 1,6-bisphosphate can cause sustained oscillations of pyruvate levels in glycolysis. In both cases, regulation mechanisms that are crucial for the switch between glycolysis and gluconeogenesis caused system failure. This agrees with earlier research on systems exhibiting highly optimized tolerance that were robust against a large range of perturbations, and yet fragile against rare perturbations[1–3].

A major part of this thesis was the development of a mathematical modelling framework to link metabolism and transcriptional regulation. This allowed us to study crosstalk between metabolism and enzyme-level regulation in amino acid metabolism in **Chapter 2**, to study the influence of transcriptional regulation on glycolysis in glycerol-producing *E. coli* in **Chapter 3**, and to perform time-course simulations of genetic knockdowns in **Chapter 4**. Here, only small models have been used to investigate the interplay between metabolism and transcriptional regulation. Small models are advantageous over larger models because they require less (often unknown) kinetic parameters. However, small models can oversimplify complex networks and neglect important interactions. Thus, in the next steps these models should be carefully extended to a larger scale. This would allow us to include the hierarchical transcription regulation network, as well as known metabolite-TF interactions[4]. Simulating CRISPRi knockdowns could then resemble the dataset in **Chapter 6** and help in developing approaches to identify new metabolite-TF interactions.

In **Chapter 6**, we used CRISPRi to knockdown 283 metabolic target genes in *E. coli* and measured proteome and metabolome in the perturbed strains. A pathway-based analysis of the proteome data then revealed transcriptional feedback regulation in 24 metabolic pathways. To systematically identify the metabolites that cause these changes, machine learning could be used. Since biological networks resemble social networks in many ways, it could be possible to develop graph-based recommender systems to predict functional interactions between metabolites and genes or proteins[5]. In contrast to purely data-driven machine learning approaches, graph-based approaches use the known network structure as prior knowledge that improves predictions[6]. For example, a similar approach was recently developed to predict polypharmacy side effects from multi omics clinical data[7]. However, supervised machine learning relies on training data, and training data is scarce with only 134 known metabolite-TF interactions. Thus, an effective link prediction algorithm to identify functional metabolite-TF interactions should be unsupervised. Moreover, the 283 conditions used in **Chapter 6** are likely not enough for machine learning applications.

Recently, the metabolome was predicted from the proteome in *Saccharomyces cerevisiae*[8]. A big question is therefore if the proteome can be predicted from the metabolome. Metabolomics can be faster and cheaper than proteomics. Moreover, if predicting the proteome from the metabolome is possible, it could also be possible to infer metabolite-TF interactions solely from metabolomics data.

Ensuring high quality metabolomics data is one of the most important prerequisites to enable data-driven inference approaches and machine learning based approaches. In the past, mass spectrometry-based proteomics has made huge progress after statistical methods, like target-decoy approaches, were developed to deal with false positive annotations[9]. Arguably, metabolomics is now at a similar stage as proteomics was back then. We showed in **Chapter 7** that FI-MS is prone to false positive annotations. The simplest solution to avoid false positive annotations in FI-MS are MS1 databases that contain peak information for several pure standards, but also for spike-in samples over many different conditions. Like in proteomics, false positives could then be detected by using database search engines[10–13].

In conclusion, understanding metabolism requires an interdisciplinary approach that combines mass spectrometry-based metabolomics, integration of multi-omics data, and mathematical modelling. Progress in these fields will enable a better understanding of *E. coli* metabolism that may even enable its rational engineering. In the future, all methods that have contributed to understanding *E. coli* could then be used to study other microbes, or even multicellular organisms.

10.1 References

(1) Csete, M. E.; Doyle, J. C. Reverse Engineering of Biological Complexity. *Science* **2002**, *295* (5560), 1664–1669. https://doi.org/10.1126/science.1069981.

(2) Carlson, J. M.; Doyle, J. Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems. *Phys. Rev. E* **1999**, *60* (2), 1412–1427. https://doi.org/10.1103/PhysRevE.60.1412.

(3) Chandra, F. A.; Buzi, G.; Doyle, J. C. Glycolytic Oscillations and Limits on Robust Efficiency. *Science* **2011**, *333* (6039), 187–192. https://doi.org/10.1126/science.1200705.

(4) Kotte, O.; Zaugg, J. B.; Heinemann, M. Bacterial Adaptation through Distributed Sensing of Metabolic Fluxes. *Mol Syst Biol* **2010**, *6*, 355. https://doi.org/10.1038/msb.2010.10.

(5) Leskovec, J.; Huttenlocher, D.; Kleinberg, J. Predicting Positive and Negative Links in Online Social Networks. In *Proceedings of the 19th international conference on World wide web - WWW '10*; ACM Press: Raleigh, North Carolina, USA, 2010; p 641. https://doi.org/10.1145/1772690.1772756.

(6) Hamilton, W. L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. 19.

(7) Zitnik, M.; Agrawal, M.; Leskovec, J. Modeling Polypharmacy Side Effects with Graph Convolutional Networks. *Bioinformatics* **2018**, *34* (13), i457–i466. https://doi.org/10.1093/bioinformatics/bty294.

(8) Zelezniak, A.; Vowinckel, J.; Capuano, F.; Messner, C. B.; Demichev, V.; Polowsky, N.; Mülleder, M.; Kamrad, S.; Klaus, B.; Keller, M. A.; Ralser, M. Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell Syst* **2018**, *7* (3), 269-283.e6. https://doi.org/10.1016/j.cels.2018.08.001.

(9) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. *Methods Mol Biol* **2010**, *604*, 55–71. https://doi.org/10.1007/978-1-60761-444-9_5.

(10) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open Mass Spectrometry Search Algorithm. *J Proteome Res* **2004**, *3* (5), 958–964. https://doi.org/10.1021/pr0499491.

(11) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am*

Soc Mass Spectrom **1994**, *5* (11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2.

(12) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

(13) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467. https://doi.org/10.1093/bioinformatics/bth092.

VI Curriculum Vitae

VII Publications of this thesis

Chapter 2:
Sander, T.; Farke, N.; Diehl, C.; Kuntz, M.; Glatter, T.; Link, H. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. Coli by Enforcing Enzyme Overabundance. Cell Syst 2019, 8 (1), 66-75. e8. https://doi.org/10.1016/j.cels.2018.12.005

Chapter 3:
Wang, C.-Y.[#]; Lempp[#], M.; Farke, N.; Donati, S.; Glatter, T.; Link, H. Metabolome and Proteome Analyses Reveal Transcriptional Misregulation in Glycolysis of Engineered E. Coli. Nat Commun 2021, 12, 4929. https://doi.org/10.1038/s41467-021-25142-0.

Chapter 4:
Donati, S.[#]; Kuntz, M.[#]; Pahl, V.; Farke, N.; Beuter, D.; Glatter, T.; Gomes-Filho, J. V.; Randau, L.; Wang, C.-Y.; Link, H. Multi-Omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms That Buffer Decreases of Enzymes in E. Coli Metabolism. *Cell Systems* **2021**, *12* (1), 56-67.e6. https://doi.org/10.1016/j.cels.2020.10.011.

Chapter 5:
Shuangyu Bi, Manika Kargeti, Remy Colin, Niklas Farke, Hannes Link, and Victor Sourjik

This chapter is part of the manuscript "Dynamic fluctuations in a bacterial metabolic network" that is currently under revision at Nature Communications

Chapter 6:
Lempp, M.; Farke, N.; Kuntz, M.; Freibert, S. A.; Lill, R.; Link, H. Systematic Identification of Metabolites Controlling Gene Expression in E. Coli. Nat Commun 2019, 10 (1), 4463. https://doi.org/10.1038/s41467-019-12474-1.

Chapter 7:
Niklas Farke[#], Thorben Schramm[#], Andreas Verhülsdonk, Hannes Link.

This chapter is part of the manuscript "A network approach identifies in-source modifications of primary metabolites during flow-injection mass spectrometry" and is currently under revision at ACS Analytical Chemistry

[#]Authors contributed equally

VIII Abgrenzung der Eigenleistung

Die Ergebnisse dieser Arbeit wurden von mir eigenständig und nur mit Hilfe der in den Kapiteln aufgeführten Personen angefertigt.

Kapitel 2 ist in Teilen veröffentlicht in Cell Systems 2019, 8 (1), 66-75. e8. Mein Anteil an dieser Studie umfasst die Entwicklung der mathematischen Modelle, die Modellanalyse, die Entwicklung der zugehörigen Grafik, sowie die Verfassung des Manuskripts.

Kapitel 3 ist in Teilen veröffentlicht in Nature Communications 2021, 12, 4929. Mein Anteil an dieser Studie umfasst die Entwicklung der mathematischen Modelle, die Entwicklung der zugehörigen Grafiken, sowie die Modellanalyse.

Kapitel 4 ist in Teilen veröffentlicht in . *Cell Systems* **2021**, *12* (1), 56-67.e6. Mein Anteil an dieser Studie umfasst die Entwicklung der mathematischen Modelle, die Entwicklung der zugehörigen Grafiken, sowie die Modellanalyse.

Kapitel 5 ist in Teil des Manuskripts „Dynamic fluctuations in a bacterial metbolic network“. Das Manuskript ist derzeit unter Revision in dem Journal Nature Communications. Mein Anteil an dieser Studie umfasst die Entwicklung des mathematischen Modells, die Entwicklung der zugehörigen Grafiken, sowie die Modellanalyse.

Kapitel 6 ist in Teilen veröffentlicht in Nat Commun 2019, 10 (1), 4463. Mein Anteil an dieser Studie umfasst die Literaturrecherche, das Design des Literatur Metabolit-TF Netzwerkes, die Entwicklung der zugehörigen Grafiken, sowie die Analyse des Netzwerkes. Weiterhin umfasst das Kapitel nicht publizierte Daten einer weiteren Studie. Mein Anteil an diesen Ergebnissen war die Datenprozessierung, die Datenanalyse, die Erstellung der Grafiken, sowie die Verfassung des Textes.

Kapitel 7 ist Teil des Manuskripts „A network approach identifies in-source modifications of primary metabolites during flow-injection mass spectrometry“. Das Manuskript ist zurzeit unter Revision bei dem Journal ACS Analytical Chemistry. Mein Anteil an dieser Studie umfasst die Durchführung der Experimente (Massenspektrometrie), die Datenprozessierung, die Datenanalyse, die Verfassung des Manuskripts, das Design der Grafiken, sowie das Design der Studie. Thorben Schramm und ich haben in gleichen Teilen zu dieser Studie beigetragen.

Kapitel 8 enthält unveröffentlichte Daten und Ergebnisse. Mein Anteil an dieser Studie war die Durchführung der Massenspektrometrie, die Datenprozessierung, die Datenanalyse, die Erstellung der Grafiken, sowie die Verfassung des Kapitels.

IX List of tables

Table 1. *Values and units of parameters and variables..... 84*
Table 2. *Literature kcat values for enzymes in amino acid biosynthesis. .....clxxxvi*
Table 3. *Amino acid requirements of E. coli .....clxxxvii*
Table 4. *Inhibition constants of allosteric enzymes..... clxxxviii*
Table 5. *CRISPRi strains and substrate metabolites and exact mass of each metabolite..... cxci*

X List of figures

Figure 1. *Central dogmatic flow of information, catabolism, and anabolism.* *15*
Figure 2. *Principles of biological robustness.* *16*
Figure 3. *Allosteric feedback and transcriptional feedback.* *19*
Figure 4. *Model-experimentation cycle..* *22*
Figure 5. *Michaelis-Menten equation.* *24*
Figure 6. *Ensemble modelling* *25*
Figure 7. *Stability and Robustness.* *28*
Figure 8. *Allosteric regulation and transcriptional regulation.* *30*
Figure 9*. Principles of flow-injection metabolomics and liquid chromatography-based tandem mass spectrometry. a) Flow-injection metabolomics (FI-MS).* *34*
Figure 10. *Thesis goals.* *37*
Figure 11*. Metabolic network and architecture of amino acid biosynthesis in E. coli..* *50*
Figure 12. *Metabolome, Proteome and fluxes of allosterically dysregulated arginine, tryptophan and histidine strains* *52*
Figure 13. *Stoichiometry and structure of the kinetic model* *53*
Figure 14. *Simulated metabolite and enzyme changes match experimental changes* *55*
Figure 15. *Mechanistic model with product inhibition* *56*
Figure 16*. Robustness-Efficiency trade-off.* *57*
Figure 17. *Enzyme levels of transcriptionally dysregulated strains.* *58*
Figure 18. *Oscillations in amino acid biosynthesis.* *59*
Figure 19. *Enzyme overabundance achieves robustness against perturbations of gene expression by CRISPR interference.* *61*
Figure 20. *Glycolysis and an engineered glycerol pathway* *73*
Figure 21*. Theoretical and experimental relationships between the glycerol production rate and the growth rate of E. coli.* *74*
Figure 22*. Stoichiometry and structure of the kinetic model* *75*
Figure 23. *Robustness analysis of three models* *76*
Figure 24*. Comparison time-course simulation and continuation method.* *78*
Figure 25*. Growth rates and glycerol titers* *79*
Figure 26*. Knockdown of CarAB causes ornithine accumulation* *93*
Figure 27. *Model structure and robustness analysis* *94*
Figure 28. *Time course simulation of the CarAB knockdown.* *95*
Figure 29. *Time course measurements of metabolite concentrations* *96*
Figure 30*. Structure and stoichiometry of a mechanistic model of glycolysis.* *107*
Figure 31. *Simulated Pyruvate oscillations* *108*
Figure 32*. Analysis of model parameter sets for their propensity to produce pyruvate* *109*
Figure 33*. Response of the metabolic model to a 10-fold upshift in glucose uptake rate* *110*
Figure 34*. Knockdown of target genes causes proteome changes* *119*
Figure 35. *KEGG pathways with putative feedback circuits..* *120*
Figure 36*. Histidine KEGG pathway* *121*
Figure 37. *Methylerythritol phosphate KEGG pathway* *122*
Figure 38. *Folate metabolism* *123*
Figure 39*. Metabolite-TF interaction map.* *124*
Figure 40. *Classification of effector metabolites into chemical groups.* *125*
Figure 41*. CRISPRi knockdowns offset the metabolome.* *126*
Figure 42*. CRISPRi knockdowns cause Substrate accumulations.* *127*
Figure 43. *Arginine and Homocysteine changes in 283 CRISPRi knockdowns.* *129*
Figure 44. *nadC and nicotinate.* *130*

Figure 45. *FI-MS with 160 authentic metabolite standards. 141*
Figure 46. *Systematic analysis of all m/z features that increase in spike-in samples..... 143*
Figure 47*. Networks of significant m/z features..... 146*
Figure 48. *Identification of in-source fragments by MS2 spectra of metabolite standards..... 148*
Figure 49*. Misannotation of significant features to metabolites..... 149*
Figure 50. *Histidine, Purine, and Isoprenoid metabolism. 161*
Figure 51. *MS1 spectra for 21 metabolites in the purine(orange), histidine (blue), and MEP pathway (green)**.** 162*
Figure 52*. Heatmap of CRISPRi knockdowns of targets in histidine, purine, and isoprenoid biosynthesis. 163*
Figure 53*. MS2 from product scans. 164*
Figure 54.*Nucleotides 13C..... clxxxix*
Figure 55. *Ion peak of reduced glutathione ..... clxxxix*
Figure 56. *Boxplot showing the relative standard deviation (RSD) of the endogenous metabolites ..... cxc*

XI List of Abbreviations

Full Name	Abbreviation
Adenosine triphosphate	ATP
Carbamoyl phosphate synthetase	CarAB
Collision induced dissociation	CID
Clustered Regularly Interspaced Short Palindromic Repeats	CRISPR
CRISPR interference	CRISPRi
Cytidine triphosphate	CTP
Adenylo succinate	DCAMP
deactivated Cas9	dCas9
dihydroxyacetone phosphate	DHAP
Escherichia coli	*E. Coli*
Ensemble Modelling for Robustness Analysis	EMRA
Electrospray Ionisation	ESI
Flux Balance Analysis	FBA
Fructose 1,6-bisphosphate	FBP
Fructose 1,6-bisphosphatase	FBPase
2-(formamido)-N-(5-phospho-b-D-ribosyl)-acetamide	FGAM
Flow-injection mass spectrometry	FI-MS
Glyceraldehyde 3-phosphate dehydrogenase	GapA
Glycerol-3-phosphate dehydrogenase 1	GPD1
Glycerol-3-phosphate phosphohydrolase 2	GPP2
Human Metabolome Database	HMDB
Kyoto Encyclopedia of Genes and Genomes	KEGG
Liquid Chromatography	LC
Liquid chromatography tandem mass spectrometry	LC-MS/MS
mass-to-charge ratio	*m/z*
Metabolic Control Analysis	MCA
Methylerythritol pathway	MEP
Metabolite and Chemical Entity Database	METLIN
Mass spectrum before fragmentation	MS1
Mass spectrum after fragmentation	MS2
Ordinary Differential Equation	ODE
Phosphoenolpyruvate	PEP
Phosphofructokinase	PFK
Phosphotransferase system	PTS
Pyruvate Kinase	PYK
Saccharomyces cerevisiae	*S. Cerevisiae*
Transcription Factor	TF
Time-of-Flight	ToF
Transcription Regulation Network	TRN
Uridine triphosphate	UTP

XII Appendix

Supplemental information related to Chapter 2

***Table 2**. Related to Figure 4; Literature kcat values for enzymes in amino acid biosynthesis. The values were collected from the BRENDA database and from Davidi and Milo, 2017[1,2]. - indicates that no value could be found in both sources. The 25th and 75th quartiles of these k_{cat} values are 930 min^{-1} and 4140 min^{-1}, respectively.*

Name	k_{cat}, s^{-1}	Name	k_{cat}, s^{-1}	Name	k_{cat}, s^{-1}
argA	654.00	*cysK*	378.50	*ilvN*	40.00
argB	-	*cysM*	24.00	*leuA*	-
argC	-	*cysN*	-	*leuB*	69.00
argD	-	*cysQ*	11.00	*leuC*	-
argE	1800.00	*dadX*	33.66	*leuD*	-
argF	-	*dapA*	104.00	*lysA*	33.00
argG	-	*dapB*	382.00	*lysC*	22.13
argH	-	*dapD*	36.00	*metA*	22.00
argI	-	*dapE*	-	*metB*	121.00
aroA	32.00	*dapF*	84.00	*metC*	34.10
aroB	14.00	*gdhA*	37.00	*metE*	3.50
aroC	39.00	*glnA*	33.00	*metH*	-
aroD	75.00	*gltB*	-	*metL*	-
aroE	237.00	*gltD*	-	*pheA*	32.00
aroF	-	*glyA*	10.00	*proA*	10.00
aroG	4.20	*hisA*	7.20	*proB*	53.00
aroH	-	*hisB*	-	*proC*	717.00
aroK	-	*hisC*	-	*prs*	-
aroL	-	*hisD*	12.00	*serA*	29.00
asd	-	*hisF*	-	*serB*	-
asnA	-	*hisG*	-	*serC*	1.80
asnB	4.50	*hisH*	-	*thrA*	-
aspC	-	*hisI*	-	*thrB*	17.00
avtA	-	*ilvA*	-	*thrC*	-
cysC	50.00	*ilvB*	38.50	*trpA*	-
cysD	-	*ilvC*	0.30	*trpB*	-
cysE	772.00	*ilvD*	69.00	*trpC*	18.77
cysH	-	*ilvE*	-	*trpE*	-
cysI	47.00	*ilvH*	-	*tyrA*	71.00
cysJ	-	*ilvI*	-	*tyrB*	-

***Table 3**. Related to Figure 4; Amino acid requirements of E. coli (Monk et al., 2017)[3]. The mean of 86.6 mM was used as parameter α in the model.*

Amino Acid	Coefficients, mmol g_{dw}^{-1}	alpha, mM
ala-L	0.499	166.4
arg-L	0.287	95.8
asn-L	0.234	78.1
asp-L	0.234	78.1
cys-L	0.089	29.7
gln-L	0.256	85.2
glu-L	0.256	85.2
gly	0.595	198.4
his-L	0.092	30.7
ile-L	0.282	94.1
leu-L	0.438	145.9
lys-L	0.333	111.1
met-L	0.149	49.8
phe-L	0.180	60.0
pro-L	0.215	71.6
ser-L	0.210	69.9
thr-L	0.247	82.2
trp-L	0.055	18.4
tyr-L	0.134	44.7
val-L	0.411	137.1
Mean	**0.260**	**86.6**

Table 4. Related to Figure 4*; Inhibition constants of allosteric enzymes (Ki-value), transcriptional attenuation (tRNA-ligase Km-value) and metabolite-transcription factor interactions (Kd-value). Values were obtained from EcoCyc[4], Brenda[1] or RegulonDB[5]. When more than one value was available, an upper and a lower bound are given. The grey background indicates the seven pathways that were investigated during this work*

Biosynthesis pathway	Allosteric Feedback		K_i mM		Transcriptional Feedback			$K_{m/d}$ mM	
	Enzyme	Metabolite	LB	UB	Mechanism	Protein	Metabolite	LB	UB
Arginine	argA	arg	0.15 [a]		Repressor	argR	arg	0.28	
Asparagine	asnA	asn	0.12		Repressor	asnC	asn	1	
Cysteine	cysE	cys	0.001						
Histidine	hisG	his	0.012	0.1	Attenuation	his-tRNA ligase	his	0.008	0.03
Isoleucine	ilvA	ile	0.06		Attenuation	ile-tRNA ligase	ile	0.0036	1.3
Leucine	leuA	leu	0.28		Attenuation	leu-tRNA ligase	leu	0.0015	0.05
Lysine	dapA	lys	0.21	3.9					
Methionine	metA	met	0.1	4	Repressor	metJ	sam	0.01	0.05
Phenylalanine	pheA	phe	0.1	0.6		tyrR	phe	>0.18	
Proline	proB	pro	0.02						
Serine	serA	ser	0.005	0.37					
Threonine	thrA	thr	0.097	0.167	Attenuation	thr-tRNA ligase	thr	0.11	0.2
Tryptophan	trpE	trp	0.17		Repressor	trpR	trp	0.16	
Tryptophan	trpE	trp	0.17		Attenuation	trp-tRNA ligase	trp	0.017	
Tyrosine	tyrA	tyr	0.1		Repressor	tyrR	tyr	0.18	
Valine	ilvB	val	0.078	0.1	Attenuation	val-tRNA ligase	val	0.0043	0.1

Supplemental information related to Chapter 7

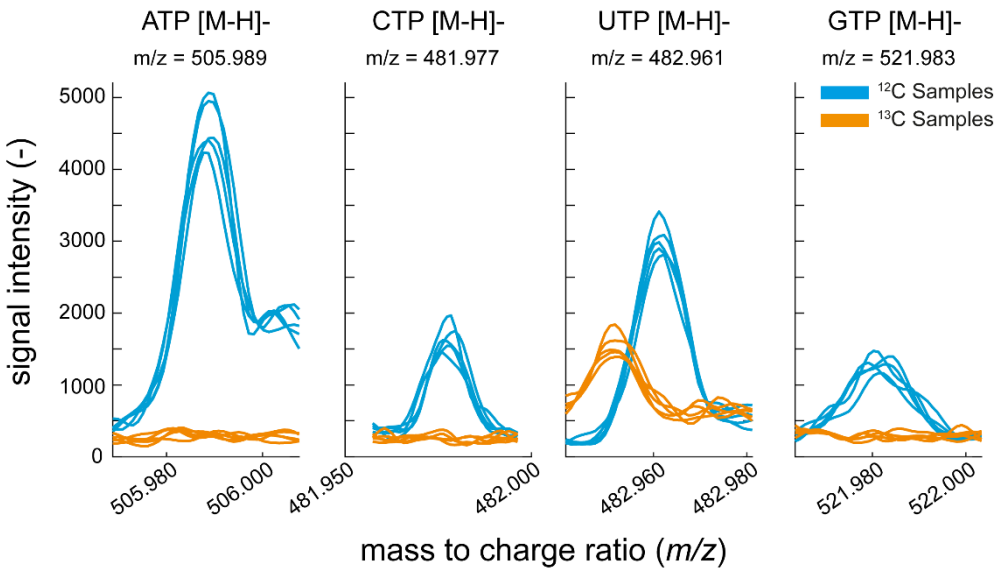


Figure 54. Ion peaks that are annotated to four nucleotides (ATP, CTP, UTP, GTP) in negative ionization mode. ^{12}C-labelled *E. coli* samples are blue lines measured in five technical replicates. ^{13}C-labelled *E. coli* samples are orange lines measured in five technical replicates. ATP, adenosine triphosphate; CTP, cytidine triphosphate; UTP, uridine triphosphate; GTP, guanosine triphosphate.

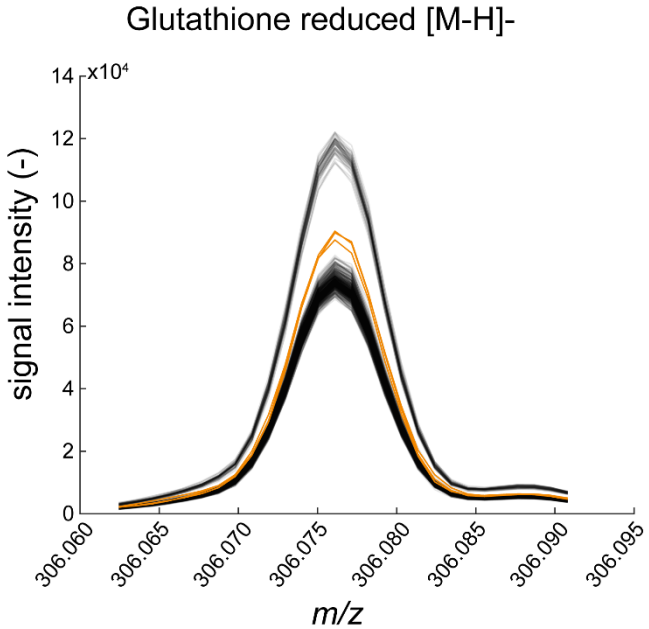


Figure 55. *Ion peak of reduced glutathione (deprotonated) in negative ionization mode. The spike-in sample that contains reduced glutathione is indicated in orange, the other 159 spike-in samples are black.*

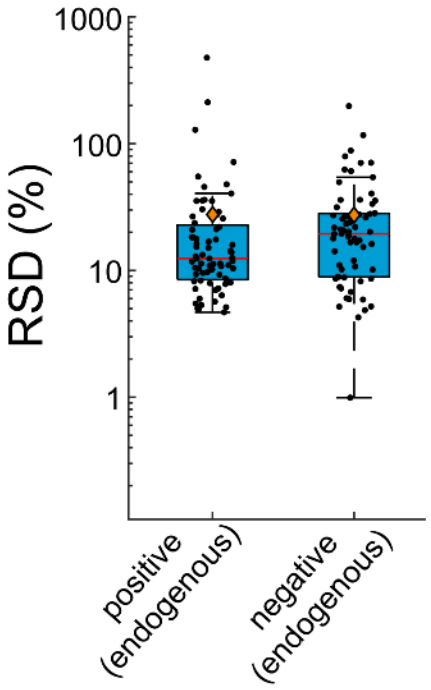


Figure 56. *Boxplot showing the relative standard deviation (RSD) of the endogenous metabolites measured by FI-MS in positive and negative ionization mode. Black dots are the RSD for each endogenous metabolite. Upper and lower box edges indicate the 25 % and 75 % percentiles. The whiskers indicate the furthest point, at which samples were not considered as outliers. The red line indicates the median. Orange diamonds are the means.*

Supplemental information related to Chapter 8

***Table 5**. CRISPRi strains and substrate metabolites and exact mass of each metabolite*

Pathway	Gene	Metabolite abbreviation	Monoisotopic mass (Da)
Histidine	*hisG*	PRPP	389.9518
Histidine	*hisI*	PRBATP	719.0043
Histidine	*hisI*	PRAMP	559.0717
Histidine	*hisA*	PRFP	577.0822
Histidine	*hisH*	PRLP	577.0822
Histidine	*hisF*	PRLP	577.0822
Histidine	*hisB*	EIG3P	238.0355
Histidine	*hisB*	HISP	221.0565
Histidine	*hisC*	IMACP	220.0249
Histidine	*hisD*	HISTD	141.0902
Histidine	*hisD*	HISTDA	139.0746
Purine	*purF*	PRPP	389.9518
Purine	*purD*	PRAM	229.0351
Purine	*purN*	GAR	286.0566
Purine	*purT*	GAR	286.0566
Purine	*purL*	FGAR	314.0515
Purine	*purM*	FGAM	313.0675
Purine	*purK*	AIR	295.0569
Purine	*purE*	N5-CAIR	339.0468
Purine	*purC*	CAIR	339.0468
Purine	*purB*	SAICAR	454.0737
Purine	*purH*	AICAR	338.0627
Purine	*purH*	FAICAR	366.0577
Purine	*purA*	IMP	348.0471
Purine	*purB*	DCAMP	463.074
MEP	*dxr*	1DXYL5P	214.0242
MEP	*ispD*	MEP	216.0399
MEP	*ispE*	CDP-ME	521.0811
MEP	*ispF*	2P4C2ME	601.0475
MEP	*ispG*	2MECDP	277.9957
MEP	*ispH*	H2MB4P	262.0007

References

(1) Schomburg, I.; Chang, A.; Schomburg, D. BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Res* **2002**, *30* (1), 47–49. https://doi.org/10.1093/nar/30.1.47.

(2) Davidi, D.; Milo, R. Lessons on Enzyme Kinetics from Quantitative Proteomics. *Current Opinion in Biotechnology* **2017**, *46*, 81–89. https://doi.org/10.1016/j.copbio.2017.02.007.

(3) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; Feist, A. M.; Palsson, B. O. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat Biotechnol* **2017**, *35* (10), 904–908. https://doi.org/10.1038/nbt.3956.

(4) Keseler, I. M.; Mackie, A.; Santos-Zavaleta, A.; Billington, R.; Bonavides-Martínez, C.; Caspi, R.; Fulcher, C.; Gama-Castro, S.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Muñiz-Rascado, L.; Ong, Q.; Paley, S.; Peralta-Gil, M.; Subhraveti, P.; Velázquez-Ramírez, D. A.; Weaver, D.; Collado-Vides, J.; Paulsen, I.; Karp, P. D. The EcoCyc Database: Reflecting New Knowledge about Escherichia Coli K-12. *Nucleic Acids Res* **2017**, *45* (D1), D543–D550. https://doi.org/10.1093/nar/gkw1003.

(5) Salgado, H.; Gama-Castro, S.; Peralta-Gil, M.; Díaz-Peredo, E.; Sánchez-Solano, F.; Santos-Zavaleta, A.; Martínez-Flores, I.; Jiménez-Jacinto, V.; Bonavides-Martínez, C.; Segura-Salazar, J.; Martínez-Antonio, A.; Collado-Vides, J. RegulonDB (Version 5.0): Escherichia Coli K-12 Transcriptional Regulatory Network, Operon Organization, and Growth Conditions. *Nucleic Acids Res* **2006**, *34* (Database issue), D394-397. https://doi.org/10.1093/nar/gkj156.